جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة الإسراء

رقم الإيداع في دار الكتب والوثائق ببغداد (2445) لسنة (2020)
الرقم الدولي للنسخة الورقية ISSN : (7145 - 2709)
للنسخة الإلكترونية E-ISSN: (7732 - 2790)

مجلة كلية الاسراء الجامعة
للعلوم الهندسية

دورية محكمة شاملة
تصدر عن جامعة الاسراء - بغداد \ العراق

المجلد السابع - العدد الحادي عشر
لسنة 2025

مجلة كلية الإسراء
الجامعة للعلوم الهندسية

مجلة علمية محكمة تصدر عن جامعة الإسراء

رقم الايداع في دار الكتب والوثائق ببغداد (2445) لسنة (2020)
الرقم الدولي للنسخة الورقية (ISSN : 2709 - 7145)
الرقم الدولي للنسخة الإلكترونية (E-ISSN: 2790-7732)

جامعة الإسراء

(سُبْحَانَ الَّذِي أَسْرَى بِعَبْدِهِ لَيْلًا مِّنَ الْمَسْجِدِ الْحَرَامِ إِلَى الْمَسْجِدِ الْأَقْصَى)

E U

AL- Esraa University

م 2011 – 1432 هـ

المجلد 7 ـ العدد 11 ـ لسنة 2025

بسم الله الرحمن الرحيم

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
دائرة البحث والتطوير

Republic of Iraq
Ministry of Higher Education & Scientific Research
Research & Development Department

الرقم: ب ت 4 / 5749
التاريخ: 2021/09/06

No.:
Date:

١٠٦٠
٢٠٢١/٩/٨

كلية الاسراء الجامعة / السيد العميد المحترم

م/ مجلة كلية الاسراء الجامعة للعلوم والهندسة

السلام عليكم ورحمة الله وبركاته ...

أشارة الى كتابكم المرقم م.ع/٢٣٩٨ في ٣١ / ١٢/ ٢٠٢٠ بشأن اعتماد مجلتكم واعتمادها لأغراض النشر والترقيات العلمية وتسجيلها ضمن موقع المجلات الاكاديمية العلمية العراقية ، حصلت موافقة السيد وكيل الوزارة لشؤون البحث العلمي بتاريخ ٢٠٢١/٨/٢٤ على أعتماد المجلة المذكورة في الترقيات العلمية والنشاطات العلمية المختلفة الأخرى ، واعتباراً من المجلد الثالث – العدد الثالث – لسنة ٢٠٢١ وتسجيل المجلة في موقع المجلات الاكاديمية العلمية العراقية.

للتفضل بالاطلاع وابلاغ مخول المجلة لمراجعة دائرتنا لتزويده بإسم المستخدم وكلمة المرور ليتسنى له تسجيل المجلة ضمن موقع المجلات العلمية العراقية وفهرسة اعدادها ... مع التقدير.

أ.م.د يوسف خلف يوسف
ء/ المدير العام لدائرة البحث والتطوير
٢٠٢١/٩/ ٦

نسخة منه الى:
- مكتب السيد وكيل الوزارة لشؤون البحث العلمي / أشارة الى موافقة سيادته المذكورة أعلاه والمثبتة على اصل مذكرتنا المرقم ب ت م٤/ ٤٥٧٦ في ٢٠٢١/٨/٢٣ / للتفضل بالاطلاع ... مع التقدير.
- قسم المشاريع الريادية / شعبة المشاريع الالكترونية / للتفضل بالعلم واتخاذ مايلزم ... مع التقدير
- قسم الشؤون العلمية / شعبة التاليف والنشر والترجمة / مع الاوليات .
- الصادرة .

مهند ابراهيم
٦ / أيلول

وزارة التعليم العالي والبحث العلمي – دائرة البحث والتطوير – القصر الأبيض – المجمع التربوي – الطابق السادس
قسم الشؤون العلمية sedep@rdd.edu.iq

تعليمات النشر

في مجلة كلية الاسراء الجامعة للعلوم الهندسية

- تصدر جامعة الاسراء (مجلة كلية الاسراء الجامعة للعلوم الهندسية) في مجلد سنوي يضم عددين.
- تقوم المجلة بنشر البحوث العلمية للباحثين في تخصصات العلوم الهندسية التالية:
  - هندسة العمارة
  - هندسة مدني
  - هندسة كيمياوية
  - هندسة الحاسوب
  - هندسة كهربائية
  - هندسة المواد
  - هندسة ميكانيكية

شروط النشر

1 - يطبع البحث بواسطة الحاسوب بمسافات مفردة بين الاسطر وبحجم خط 12 ونوع (Simplified Arabic)، اما العنوان باللغتين العربية والانكليزية فيكون بحجم خط 14 شريطة الا يزيد عدد صفحاته عن 15 صفحة بما في ذلك الجداول والاشكال والمراجع وعلى وجه واحد على ورق قياس A4 مع ترك هامش في حدود 2 سم من الاعلى والاسفل وهامش بحدود 3 سم من الجانبين الايمن والايسر.

2 - لا يفضل نشر البحوث من قبل رئيس واعضاء هيئة التحرير في المجلة سواء كان البحث منفرداً أو مشتركاً.

3 - يقدم البحث بثلاث نسخ ورقية ونسخة الكترونية بعد قبول البحث للنشر، يسلم البحث بشكله النهائي مطبوعا بالنظام الاعتيادي بمسافة منتظمة لكافة الصفحات عدا الصفحة الاولى التي تتضمن عنوان البحث و اسماء الباحثين وعناوينهم باللغتين العربية والإنكليزية متبوعا بالبريد الالكتروني للباحث الاول وعلى قرص مرن CD ببرنامج Microsoft Word / 2010.

4 - تقبل البحوث باللغتين العربية والانكليزية ويفضل كتابة البحث باللغة الانكليزية.

دليل المؤلف Author Guidelines

ادناه الشروط والمتطلبات الواجب مراعاتها من قبل الباحث للنشر في هذه المجلة بشرط أن لا يكون البحث قد نشر أو سينشر في أية مجلة هندسية أخرى ولم يمضِ على انجازه اكثر من أربع سنوات.

1 - يجب ان يكون عنوان البحث موجزاً قدر الامكان ومعبر عن البحث.

2 - اسماء الباحثين: تكتب اسماء الباحثين وعناوين عملهم بصورة واضحة مع البريد الالكتروني للباحث الاول.

3 - يجب ان يتضمن المستخلص موجزا واضحا عن البحث مكون من 250-300 كلمة متبوعا بكلمات مفتاحية 4-6. أذا كان البحث باللغة العربية فيكون المستخلص متبوعا بالكلمات المفتاحية اولا ثم المستخلص متبوعا بالكلمات المفتاحية باللغة الانكليزية ثانيا و العكس صحيح.

4 - المقدمة: تتضمن مراجعة المعلومات وثيقة الصلة بموضوع البحث الموجودة في المصادر العلمية وتنتهي المقدمة باهداف الدراسة وأساسها المنطقي.

5 - المواد وطرائق العمل: تذكر طرائق العمل بشكل مفصل ان كانت جديدة اما اذا كانت منشورة فتذكر بشكل مختصر مع الاشارة للمصدر وتستعمل وحدات النظام العالميSystem International of Units (S.I.Us)

6 - النتائج والمناقشة: تعرض بشكل موجز وهادف وبنظام متوالي وتعرض النتائج بافضل صورة معبرة وتوضع الجداول والاشكال في أماكنها المخصصة بعد الاشارة إليها في النتائج.

7 - يستعمل نظام الارقام العربية وهكذا في البحوث المرسلة للنشر وتمثل مناقشة النتائج تعبيراً موجزاً عن النتائج وتفسيراتها.

8 - تكون كتابة المصدر في القائمة المصادر متضمنة الآتي: اسم او أسماء الباحثين، سنة النشر وعنوان البحث كاملاً واسم المجلة ورقم المجلد والعدد وعدد الصفحات، مثال:

الخفاجي، جاسم محمود و حميد، محمد حسوني و كريم، حيدر حاتم، (2018) " دراسة تجريبية على الخرسانة مع استبدال جزئي للركام الخشن بواسطة المطاط غير المرغوب فيه".

مجلة كلية الاسراء الجامعة، المجلد 1 العدد 1، 217-243. و ممكن ان تكتب كالاتي: مجلة كلية الاسراء الجامعة، 1(1)،217-243.

9 - المستخلص الانكليزي يجب أن يكون وافياً ومعبراً عن البحث بصورة دقيقة وليس بالضرورة ان يكون ترجمة حرفية للمستخلص العربي و متبوعا بكلمات مفتاحية 4-6.

دليل المقيّم Reviewer Guidelines

أدناه الشروط والمتطلبات الواجب مراعاتها من قبل المقيم للبحوث المرسلة للنشر في هذه المجلة

1 - ملأ استمارة التقويم المرسلة رفقة البحث المطلوب تقييمه بشكل دقيق وعدم ترك أي فقرة بدون اجابة.

2 - على المقيّم التاكد من تطابق وتوافق عنوان البحث باللغتين العربية والانكليزية وفي حالة عدم تطابقهما اقتراح العنوان البديل.

3 - أن يبين المقيّم هل ان الجداول والاشكال التخطيطية الموجودة في البحث وافية ومعبرة.

4 - أن يبين المقيّم هل ان الباحث اتبع الاسلوب الإحصائي الصحيح.
5 - أن يوضح المقيّم هل ان مناقشة النتائج كانت كافية ومنطقية.
6 - على المقيّم تحديد مدى استخدام الباحث للمراجع العلمية الرصينة وحداثتها.
7 - أن يؤشر المقيّم بشكل واضح على واحد من ثلاث اختيارات وهي:
البحث صالح للنشر بدون تعديلات.
البحث صالح للنشر بعد اجراء التعديلات.
البحث غير صالح للنشر.
8 - يجب أن يوضح المقيّم بورقة منفصلة ما هي التعديلات الأساسية التي يقترحها لغرض قبول البحث.
9 - للمقيّم حق طلب إعادة البحث إليه بعد إجراء التعديلات المطلوبة للتأكد من التزام الباحث بها.
10 -على المقيّم تسجيل اسمه ودرجته العلمية وعنوانه وتاريخ اجراء التقييم مع التوقيع على استمارة التقييم المرسلة له رفقه البحث المرسل له للتقييم.

المصادر

1 - يشار الى المصادر في متن البحث كما يلي:
اللقب او الاسم الثالث للمؤلف والسنة اذا كان البحث بإسم باحث واحد، واذا كان مؤلفين فيذكران والسنة واذا كانوا ثلاثة فاكثر فيذكر اسم الاول واخرون والسنة.
2 - ترتب المصادر حسب الصيغة العالمية (APA) وكما بالامثلة المذكورة:
أ- بحث في مجلة.
اسم الباحث أو الباحثون، (السنة)، عنوان البحث، اسم المجلة، المجلد، العدد و صفحتي البدء والانتهاء للبحث.
ب- كتب.
اسم المؤلف أو المؤلفون، (السنة) عنوان الكتاب، الطبعة، دار النشر وعدد الصفحات.

ج- الرسائل والاطاريح الجامعية.

اسم الباحث، (السنة)، عنوان الرسالة او الاطروحة، العنوان (الكلية والجامعة) وعدد الصفحات.

د- بحث في وقائع مؤتمر او ندوة علمية.

اسم الباحث أو الباحثون، (السنة)، عنوان البحث، اسم المؤتمر او الندوة العلمية، مكان الانعقاد، صفحتي البدء والانتهاء للبحث.

ترسل البحوث الى مجلة كلية الاسراء الجامعة للعلوم الهندسية على العنوان الاتي:

جامعة الاسراء– قسم التوثيق والنشر

بغداد / العراق

البريد الالكتروني:

al-esraajournal@esraa.edu.iq

(تعهد الملكية الفكرية)

إني الباحث..... صاحب البحث الموسوم (.....
.....)

أتعهد بأن البحث قد أنجز من قبلي ولم ينشر في مجلة أخرى في داخل وخارج العراق وأرغب بنشره في مجلة (**مجلة كلية الإسراء الجامعة للعلوم الهندسية**) التي تصدرها جامعة الإسراء.

التوقيع:

التاريخ:

(تعهد نقل حقوق الطبع والتوزيع)

إني الباحث..... صاحب البحث الموسوم (.....
.....)

أتعهد بنقل حقوق الطبع والتوزيع والنشر إلى مجلة (**مجلة كلية الإسراء الجامعة للعلوم الهندسية**) التي تصدرها جامعة الإسراء.

التوقيع:

التاريخ:

المحتويات

[137]. *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 2500–2504. doi: 10.1109/ICASSP39728.2021.9414582.

[138]. F. Marra, C. Saltori, G. Boato, and L. Verdoliva, ―Incremental learning for the detection and classification of gan-generated images,‖ in *2019 IEEE international workshop on information forensics and security (WIFS)*, 2019, pp. 1–6.

[139]. A. Bharati, R. Singh, M. Vatsa, and K. W. Bowyer, ―Detecting Facial Retouching Using Supervised Deep Learning,‖ *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 9, pp. 1903–1913, 2016, doi: 10.1109/TIFS.2016.2561898.

[140]. C. Kong, B. Chen, H. Li, S. Wang, A. Rocha, and S. T. W. Kwong, ―Detect and Locate: A Face Anti-Manipulation Approach with Semantic and Noise-level Supervision,‖ *ArXiv*, vol. abs/2107.0, 2021.

[141]. L. Cao, W. Sheng, F. Zhang, K. Du, C. Fu, and P. Song, ―Face Manipulation Detection Based on Supervised Multi-Feature Fusion Attention Network.,‖ *Sensors (Basel).*, vol. 21, no. 24, Dec. 2021, doi: 10.3390/s21248181.

[142]. I. Amerini, L. Galteri, R. Caldelli, and A. Del Bimbo, ―Deepfake video detection through optical flow based cnn,‖ in

[143]. *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 2019, p. 0.

[144]. T. Jung, S. Kim, and K. Kim, ―Deepvision: Deepfakes detection using human eye blinking pattern,‖ *IEEE Access*, vol. 8, pp. 83144–83154, 2020.

[126]. M. Barni, K. Kallas, E. Nowroozi, and B. Tondi, ―CNN Detection of GAN-Generated Face Images based on Cross-Band Co- occurrences Analysis,‖ in *2020 IEEE International Workshop on Information Forensics and Security (WIFS)*, IEEE, Dec. 2020, pp. 1–6. doi: 10.1109/WIFS49906.2020.9360905.

[127]. H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. K. Jain, ―On the Detection of Digital Face Manipulation,‖ in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5780–5789. doi: 10.1109/CVPR42600.2020.00582.

[128]. R. Wang, L. Ma, F. Juefei-Xu, X. Xie, J. Wang, and Y. Liu, ―FakeSpotter: {A} Simple Baseline for Spotting AI-Synthesized Fake Faces,‖ *CoRR*, vol. abs/1909.0, 2019, [Online]. Available: http://arxiv.org/abs/1909.06122

[129]. A. Jain, P. Majumdar, R. Singh, and M. Vatsa, ―Detecting GANs and Retouching based Digital Alterations via DAD-HCNN,‖ in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, pp. 2870–2879. doi: 10.1109/CVPRW50498.2020.00344.

[130]. Z. Mi, X. Jiang, T. Sun, and K. Xu, ―GAN-Generated Image Detection With Self-Attention Mechanism Against GAN

[131]. Generator Defect,‖ *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 5, pp. 969–981, 2020, doi: 10.1109/JSTSP.2020.2994523.

[132]. X. Zhang, S. Karaman, and S.-F. Chang, ―Detecting and Simulating Artifacts in GAN Fake Images,‖ in *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*, IEEE, Dec. 2019, pp. 1–6. doi: 10.1109/WIFS47025.2019.9035107.

[133]. T. Dzanic, K. Shah, and F. D. Witherden, ―Fourier Spectrum Discrepancies in Deep Network Generated Images,‖ in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, in NIPS'20. Red Hook, NY, USA: Curran Associates Inc., 2020.

[134]. R. Durall, M. Keuper, and J. Keuper, ―Watch your up-convolution: CNN based generative deep neural networks are failing to reproduce spectral distributions,‖ in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7887–7896. doi: 10.1109/CVPR42600.2020.00791.

[135]. N. Bonettini, P. Bestagini, S. Milani, and S. Tubaro, ―On the use of Benford's law to detect GAN-generated images,‖ in *2020 25th International Conference on Pattern Recognition (ICPR)*, IEEE, Jan. 2021, pp. 5495–5502. doi: 10.1109/ICPR48806.2021.9412944.

[136]. S. Hu, Y. Li, and S. Lyu, ―Exposing GAN-Generated Faces Using Inconsistent Corneal Specular Highlights,‖ in *ICASSP 2021*

[111]. Z. A. Salih, R. Thabit, K. A. Zidan, and B. E. Khoo, ―A new face image manipulation reveal scheme based on face detection and image watermarking,‖ in *2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, IEEE, 2022, pp. 1–6.

[112]. Z. A. Salih, R. Thabit, and K. A. Zidan, ―A New Manipulation Detection and Localization Scheme,‖ *J. Eng. Sci. Technol.*, vol. 18, no. 2, pp. 1164–1183, 2023.

[113]. M. H. Al-Hadaad, R. Thabit, and K. A. Zidan, ―A New Face Image Authentication Scheme based on Bicubic Interpolation,‖

[114]. *Al-Iraqia J. Sci. Eng. Res.*, vol. 2, no. 2, pp. 1000–1006, Jun. 2023, doi: 10.58564/IJSER.2.2.2023.68.

[115]. M. H. Al-Hadaad, R. Thabit, and K. A. Zidan, ―A New Face Region Recovery Algorithm based on Bicubic Interpolation,‖

[116]. *JOIV Int. J. Informatics Vis.*, vol. 7, no. 3, pp. 1000–1006, Sep. 2023, doi: 10.30630/joiv.7.3.1671.

[117]. M. H. Al-Hadaad, R. Thabit, and K. A. Zidan, ―Tamper detection , localization , and recovery for digital face images,‖ 2020.

[118]. S. Hu, Y. Li, and S. Lyu, ―Exposing GAN-Generated Faces Using Inconsistent Corneal Specular Highlights,‖ in *ICASSP 2021*

[119]. *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Jun. 2021, pp. 2500– 2504. doi: 10.1109/ICASSP39728.2021.9414582.

[120]. X. Han, Z. Ji, and W. Wang, ―Low Resolution Facial Manipulation Detection,‖ in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 2020, pp. 431–434. doi: 10.1109/VCIP49819.2020.9301796.

[121]. X. Yang, Y. Li, H. Qi, and S. Lyu, ―Exposing GAN-synthesized Faces Using Landmark Locations,‖ in *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, New York, NY, USA: ACM, Jul. 2019, pp. 113–118. doi: 10.1145/3335203.3335724.

[122]. S. McCloskey and M. Albright, ―Detecting GAN-Generated Imagery Using Saturation Cues,‖ in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 4584–4588. doi: 10.1109/ICIP.2019.8803661.

[123]. H. Li, B. Li, S. Tan, and J. Huang, ―Detection of Deep Network Generated Images Using Disparities in Color Components,‖

[124]. *ArXiv*, vol. abs/1808.0, 2018.

[125]. L. Nataraj *et al.*, ―Detecting GAN generated Fake Images using Co-occurrence Matrices,‖ *ArXiv*, vol. abs/1903.0, 2019.

[99]. A. Jain, R. Singh, and M. Vatsa, ―On Detecting GANs and Retouching based Synthetic Alterations,‖ in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, IEEE, Oct. 2018, pp. 1–7. doi: 10.1109/BTAS.2018.8698545.

[100]. R. Raghavendra, K. B. Raja, and C. Busch, ―Detecting morphed face images,‖ in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, IEEE, Sep. 2016, pp. 1–7. doi: 10.1109/BTAS.2016.7791169.

[101]. C. Rathgeb, C.-I. Satnoianu, N. E. Haryanto, K. Bernardo, and C. Busch, ―Differential Detection of Facial Retouching: A Multi-Biometric Approach,‖ *IEEE Access*, vol. 8, pp. 106373–106385, 2020, doi: 10.1109/ACCESS.2020.3000254.

[102]. U. Scherhag, D. Budhrani, M. Gomez-Barrero, and C. Busch, *Detecting morphed face images using facial landmarks*, vol. 10884 LNCS. Springer International Publishing, 2018. doi: 10.1007/978-3-319-94211-7_48.

[103]. U. Scherhag, C. Rathgeb, J. Merkle, and C. Busch, ―Deep Face Representations for Differential Morphing Attack Detection,‖

[104]. *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 3625–3639, 2020, doi: 10.1109/TIFS.2020.2994750.

[105]. S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. Efros, ―Detecting Photoshopped Faces by Scripting Photoshop,‖ in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2019, pp. 10071–10080. doi: 10.1109/ICCV.2019.01017.

[106]. U. Scherhag, J. Kunze, C. Rathgeb, and C. Busch, ―Face morph detection for unknown morphing algorithms and image sources: a multi-scale block local binary pattern fusion approach,‖ *IET Biometrics*, vol. 9, no. 6, pp. 278--289(11), Nov. 2020, [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2019.0206

[107]. C. Rathgeb *et al.*, ―PRNU-based detection of facial retouching,‖ *IET Biometrics*, vol. 9, no. 4, pp. 154--164(10), Jul. 2020, [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2019.0196

[108]. F. Marra, D. Gragnaniello, D. Cozzolino, and L. Verdoliva, ―Detection of GAN-Generated Fake Images over Social Networks,‖ in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2018, pp. 384–389. doi: 10.1109/MIPR.2018.00084.

[109]. D. Gragnaniello, D. Cozzolino, F. Marra, G. Poggi, and L. Verdoliva, ―Are GAN Generated Images Easy to Detect? A Critical Analysis of the State-Of-The-Art,‖ in *2021 IEEE International Conference on Multimedia and Expo (ICME)*, 2021, pp. 1–6. doi: 10.1109/ICME51207.2021.9428429.

[110]. T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi, ―Deep Learning for Deepfakes Creation and Detection,‖ *CoRR*, vol. abs/1909.1, 2019, [Online]. Available: http://arxiv.org/abs/1909.11573

[86]. E. Fourati, W. Elloumi, and A. Chetouani, ―Anti-spoofing in face recognition-based biometric authentication using Image Quality Assessment,‖ *Multimed. Tools Appl.*, vol. 79, pp. 865–889, 2019.

[87]. L. Li, P. L. Correia, and A. Hadid, ―Face recognition under spoofing attacks: countermeasures and research directions,‖ *IET Biom.*, vol. 7, pp. 3–14, 2018.

[88]. A. Snook, ―What is Deepfake Identity Theft?,‖ *i-Sight website*. 2020. [Online]. Available: https://www.i- sight.com/resources/what-is-deepfake-identity-theft/

[89]. E. Haller, ―The two faces of deepfakes: Cybersecurity & identity fraud,‖ *Secur. Mag.*, 2022, [Online]. Available: https://www.securitymagazine.com/articles/97085-the-two-faces-of-deepfakes-cybersecurity-and-identity-fraud

[90]. R. Hendrikse, ―How Deepfakes Could Become A Threat To Your Identity,‖ *Forbes*, 2019, [Online]. Available: https://www.forbes.com/sites/renehendrikse/2019/12/20/how-deepfakes-could-become-a-threat-to-your- identity/?sh=3ce7083e1063

[91]. L. Patel, ―The Rise of Deepfakes and What That Means for Identity Fraud,‖ *DarkReading Authentication*, 2020, [Online].

[92]. Available: https://www.darkreading.com/authentication/the-rise-of-deepfakes-and-what-that-means-for-identity-fraud

[93]. J. Cote, ―DEEPFAKES AND FAKE NEWS POSE A GROWING THREAT TO DEMOCRACY, EXPERTS WARN,‖ *News Northeast.*, 2022, [Online]. Available: https://news.northeastern.edu/2022/04/01/deepfakes-fake-news-threat-democracy/

[94]. V. Vieira, ―Deepfakes and the intensification of fake news,‖ *Inst. Res. internet Soc.*, 2019, [Online]. Available: https://irisbh.com.br/en/deepfakes-and-the-intensification-of-fake-news/

[95]. C. Vaccari and A. Chadwick, ―Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News,‖ *Soc. Media + Soc.*, vol. 6, no. 1, p. 2056305120903408, Jan. 2020, doi: 10.1177/2056305120903408.

[96]. Z. Akhtar, D. Dasgupta, and B. Banerjee, ―Face Authenticity: An Overview of Face Manipulation Generation, Detection and Recognition,‖ *SSRN Electron. J.*, 2019.

[97]. U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch, ―Face Recognition Systems Under Morphing Attacks: A Survey,‖ *IEEE Access*, vol. 7, pp. 23012–23026, 2019, doi: 10.1109/ACCESS.2019.2899367.

[98]. J. Frank, T. Eisenhofer, L. Schönherr, A. Fischer, D. Kolossa, and T. Holz, ―Leveraging frequency analysis for deep fake image recognition,‖ in *37th International Conference on Machine Learning, ICML 2020*, in ICML'20, vol. PartF16814. JMLR.org, 2020, pp. 3205–3216.

[70]. J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Niessner, ―Face2Face: Real-Time Face Capture and Reenactment of RGB Videos,‖ in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 2387–2395. doi: 10.1109/ CVPR.2016.262.

[71]. R. Salia, ―Top 10 Best Alternatives to Deep Nostalgia App,‖ *Top Ten AI website*. 2021.

[72]. P. Panyatham, ―Deepfake Technology In The Entertainment Industry: Potential Limitations And Protections,‖ *Emerg. Technol.*, vol. 3, no. 10, 2020.

[73]. V. Caron and V. Bergeron, ―WHAT ARE DEEPFAKE'S IMPACTS ON THE MEDIA AND ENTERTAINMENT INDUSTRY?,‖ *CANADA MEDIA FUND*, 2019.

[74]. J. ALDREDGE, ―Is Deepfake Technology the Future of the Film Industry?,‖ *Prem. Beat by shutterstock*, 2020.

[75]. C. Rathgeb, A. Dantcheva, and C. Busch, ―Impact and Detection of Facial Beautification in Face Recognition: An Overview,‖

[76]. *IEEE Access*, vol. 7, pp. 152667–152678, 2019, doi: 10.1109/ACCESS.2019.2948526.

[77]. E. Sarkar, P. Korshunov, L. Colbois, and S. Marcel, ―Vulnerability Analysis of Face Morphing Attacks from Landmarks and Generative Adversarial Networks,‖ *ArXiv*, vol. abs/2012.0, 2020.

[78]. C. Gray, ―Add to Cart: Why deepfakes are good for retail,‖ *AdNews Newsl.*, 2020.

[79]. H. Kronk, ―UDACITY DEVELOPED AN AI TO TURN INSTRUCTOR'S SPEECH INTO DEEPFAKE-STYLE VIDEO, BUT HAS NO PLANS TO USE IT,‖ *Elearning Insid.*, 2019.

[80]. P. Korshunov and S. Marcel, ―Vulnerability of Face Recognition to Deep Morphing,‖ *ArXiv*, Oct. 2019, [Online]. Available: http://arxiv.org/abs/1910.01933

[81]. U. Anchalia, K. P. Reddy, A. Modi, K. Neelam, D. Prasad, and V. Nath, ―Study and Design of Biometric Security Systems: Fingerprint and Speech Technology,‖ in *Lecture Notes in Electrical Engineering*, 2019, pp. 577–584. doi: 10.1007/978-981- 13-7091-5_47.

[82]. A. Kamboj, R. Rani, and A. Nigam, ―A comprehensive survey and deep learning-based approach for human recognition using ear biometric,‖ *Vis. Comput.*, pp. 1–34, 2021.

[83]. U. Muhammad, T. Holmberg, W. C. de Melo, and A. Hadid, ―Face Anti-Spoofing via Sample Learning Based Recurrent Neural Network (RNN),‖ in *BMVC*, 2019.

[84]. G. GencyV, M. K. Chaithanya, and A. F. Majeed, ―Face Spoofing Detection: A Survey on Different Methodologies,‖ 2020.

[85]. Z. Boulkenafet, Z. Akhtar, X. Feng, and A. Hadid, ―Face Anti-spoofing in Biometric Systems,‖ 2017.

[57]. S. Venkatesh, R. Ramachandra, K. Raja, and C. Busch, —Face Morphing Attack Generation and Detection: A Comprehensive Survey,‖ *IEEE Trans. Technol. Soc.*, vol. 2, no. 3, pp. 128–145, Sep. 2021, doi: 10.1109/TTS.2021.3066254.

[58]. Y. Weng, L. Wang, X. Li, M. Chai, and K. Zhou, —Hair Interpolation for Portrait Morphing,‖ *Comput. Graph. Forum*, vol. 32, pp. 79–84, 2013.

[59]. H. Zhang, S. K. Venkatesh, R. Ramachandra, K. B. Raja, N. Damer, and C. Busch, —MIPGAN—Generating Strong and High Quality Morphing Attacks Using Identity Prior Driven GAN,‖ *IEEE Trans. Biometrics, Behav. Identity Sci.*, vol. 3, pp. 365– 383, 2021.

[60]. E. Gonzalez-Sosa, J. Fierrez, R. Vera-Rodriguez, and F. Alonso-Fernandez, —Facial Soft Biometrics for Recognition in the Wild: Recent Works, Annotation, and COTS Evaluation,‖ *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 8, pp. 2001–2014, Aug. 2018, doi: 10.1109/TIFS.2018.2807791.

[61]. Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, —StarGAN: Unified Generative Adversarial Networks for Multi- domain Image-to-Image Translation,‖ in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 8789–8797. doi: 10.1109/CVPR.2018.00916.

[62]. A. Agarwal, R. Singh, M. Vatsa, and A. Noore, —SWAPPED! Digital face presentation attack detection via weighted local magnitude pattern,‖ in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE, Oct. 2017, pp. 659–665. doi: 10.1109/BTAS.2017.8272754.

[63]. I. Snap, —Snapchat,‖ *App Store Preview*. 2022.

[64]. L. LOIC, —The 9 Best AI Video Generators (Text-to-Video),‖ *Make Use of Website*. 2021.

[65]. O. Fried *et al.*, —Text-based editing of talking-head video,‖ *ACM Trans. Graph.*, vol. 38, no. 4, pp. 1–14, Aug. 2019, doi: 10.1145/3306346.3323028.

[66]. F. T. Support, —Talking Avatar,‖ *Talking avatar website*. 2022.

[67]. R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, —Deepfakes and beyond: A Survey of face manipulation and fake detection,‖ *Inf. Fusion*, vol. 64, pp. 131–148, Dec. 2020, doi: 10.1016/j.inffus.2020.06.014.

[68]. H. Allcott and M. Gentzkow, —Social Media and Fake News in the 2016 Election,‖ *J. Econ. Perspect.*, vol. 31, no. 2, pp. 211– 236, May 2017, doi: 10.1257/jep.31.2.211.

[69]. D. Boneh, A. J. Grotto, P. McDaniel, and Ni. Papernot, —How Relevant Is the Turing Test in the Age of Sophisbots?,‖ *IEEE Secur. Priv.*, vol. 17, no. 6, pp. 64–71, Nov. 2019, doi: 10.1109/MSEC.2019.2934193.

[42]. P. Yu, Z. Xia, J. Fei, and Y. Lu, —A Survey on Deepfake Video Detection,‖ *IET Biometrics*, vol. 10, no. 6, pp. 607–624, Nov. 2021, doi: 10.1049/bme2.12031.

[43]. A. M. Almars, —Deepfakes Detection Techniques Using Deep Learning: A Survey,‖ *J. Comput. Commun.*, vol. 09, no. 05, pp. 20–35, 2021, doi: 10.4236/jcc.2021.95003.

[44]. L. A. Passos, D. Jodas, K. A. P. da Costa, L. A. S. Júnior, D. Colombo, and J. P. Papa, —A Review of Deep Learning-based Approaches for Deepfake Content Detection,‖ *arXiv:2202.06095v1*. arXiv, 2022. [Online]. Available: http://arxiv.org/abs/2202.06095

[45]. F. Juefei-Xu, R. Wang, Y. Huang, Q. Guo, L. Ma, and Y. Liu, —Countering malicious deepfakes: Survey, battleground, and horizon,‖ *Int. J. Comput. Vis.*, vol. 130, no. 7, pp. 1678–1734, 2022.

[46]. A. Malik, M. Kuribayashi, S. M. Abdullahi, and A. N. Khan, —DeepFake Detection for Human Face Images and Videos: A Survey,‖ *IEEE Access*, vol. 10, pp. 18757–18775, 2022, doi: 10.1109/ACCESS.2022.3151186.

[47]. A. D. and S. B. Wankhade, *Intelligent Computing and Networking*, vol. 146. 2021. [Online]. Available: http://link.springer.com/10.1007/978-981-15-7421-4

[48]. M. Ibsen, C. Rathgeb, D. Fischer, P. Drozdowski, and C. Busch, —Digital Face Manipulation in Biometric Systems,‖ C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, and C. Busch, Eds., Cham: Springer International Publishing, 2022, pp. 27–43. doi: 10.1007/978-3-030-87664-7_2.

[49]. I. Papastratis, —Deepfakes: Face synthesis with GANs and Autoencoders,‖ *AI Summer*, 2020, [Online]. Available: https://theaisummer.com/deepfakes/

[50]. D. Siegel, C. Kraetzer, S. Seidlitz, and J. Dittmann, —Media Forensics Considerations on DeepFake Detection with Hand- Crafted Features,‖ *J. Imaging*, vol. 7, no. 7, p. 108, Jul. 2021, doi: 10.3390/jimaging7070108.

[51]. F. Matern, C. Riess, and M. Stamminger, —Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations,‖ in *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, IEEE, Jan. 2019, pp. 83–92. doi: 10.1109/WACVW.2019.00020.

[52]. M. Kowalski, —FaceSwap,‖ *GetHub official website*. 2021. [Online]. Available: https://github.com/MarekKowalski/FaceSwap

[53]. A. Store, —ZAO,‖ *Changsha Shenduronghe Network Technology Co., Ltd.* 2019. [Online]. Available: https://apps.apple.com/cn/app/id1465199127

[54]. A. Verma, —9 Best Photo Morph Apps for Android & iOS in 2022,‖ *The Unfolder*, 2022.

[55]. [54] G. Wolberg, —Image morphing: a survey,‖ *Vis. Comput.*, vol. 14, no. 8, pp. 360–372, 1998, doi: 10.1007/s003710050148.

[56]. M. Gomez-Barrero, C. Rathgeb, U. Scherhag, and C. Busch, —Is your biometric system robust to morphing attacks?,‖ in *2017 5th International Workshop on Biometrics and Forensics (IWBF)*, IEEE, Apr. 2017, pp. 1–6. doi: 10.1109/IWBF.2017.7935079.

[29]. A. Rocha, W. Scheirer, T. Boult, and S. Goldenstein, —Vision of the Unseen: Current Trends and Challenges in Digital Image and Video Forensics,‖ *ACM Comput. Surv.*, vol. 43, no. 4, Oct. 2011, doi: 10.1145/1978802.1978805.

[30]. M. C. Stamm and K. J. R. Liu, Forensic detection of image manipulation using statistical intrinsic fingerprints,‖ *IEEE Trans. Inf. Forensics Secur.*, vol. 5, no. 3, pp. 492–506, 2010, doi: 10.1109/TIFS.2010.2053202.

[31]. A. Swaminathan, M. Wu, and K. J. R. Liu, —Digital image forensics via intrinsic fingerprints,‖ *IEEE Trans. Inf. Forensics Secur.*, vol. 3, no. 1, pp. 101–117, Mar. 2008, doi: 10.1109/TIFS.2007.916010.

[32]. D. Cozzolino, A. Rossler, J. Thies, M. Niesner, and L. Verdoliva, —ID-Reveal: Identity-aware DeepFake Video Detection,‖ in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2021, pp. 15088–15097. doi: 10.1109/ICCV48922.2021.01483.

[33]. A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, —FaceForensics++: Learning to Detect Manipulated Facial Images,‖ in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2019, pp. 1–11. doi: 10.1109/ICCV.2019.00009.

[34]. M. Dang and T. N. Nguyen, —Digital Face Manipulation Creation and Detection: A Systematic Review,‖ *Electronics*, vol. 12, no. 16, p. 3407, Aug. 2023, doi: 10.3390/electronics12163407.

[35]. X. Ju, —An Overview of Face Manipulation Detection,‖ *J. Cyber Secur.*, vol. 2, no. 4, pp. 197–207, 2020, doi: 10.32604/jcs.2020.014310.

[36]. R. Thakur and R. Rohilla, —Recent advances in digital image manipulation detection techniques: A brief review,‖ *Forensic Sci. Int.*, vol. 312, p. 110311, 2020.

[37]. L. Verdoliva, —Media Forensics and DeepFakes: An Overview,‖ *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 5, pp. 910– 932, Aug. 2020, doi: 10.1109/JSTSP.2020.3002101.

[38]. M. Abdolahnejad and P. X. Liu, —Deep learning for face image synthesis and semantic manipulations: a review and future perspectives,‖ *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5847–5880, Dec. 2020, doi: 10.1007/s10462-020-09835-4.

[39]. X. Zheng, Y. Guo, H. Huang, Y. Li, and R. He, —A Survey of Deep Facial Attribute Analysis,‖ *Int. J. Comput. Vis.*, vol. 128, no. 8–9, pp. 2002–2034, Sep. 2020, doi: 10.1007/s11263-020-01308-z.

[40]. S. Pashine, S. Mandiya, P. Gupta, and R. Sheikh, —Deep Fake Detection: Survey of Facial Manipulation Detection Solutions,‖ Int. Res. J. Eng. Technol., vol. 8, no. 8, pp. 4441–4449, 2021, [Online]. Available: http://arxiv.org/abs/2106.12605

[41]. Y. Mirsky and W. Lee, —The creation and detection of deepfakes: A survey,‖ *ACM Comput. Surv.*, vol. 54, no. 1, pp. 1–41, 2021.

[14]. L. Verdoliva, ―Media Forensics and DeepFakes: An Overview,‖ *IEEE J. Sel. Top. Signal Process.*, vol. 14, pp. 910–932, 2020.

[15]. A. Czajka, W. Kasprzak, and A. Wilkowski, ―Verification of iris image authenticity using fragile watermarking,‖ *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 64, no. 4, pp. 807–819, Dec. 2016, doi: 10.1515/bpasts-2016-0090.

[16]. I. J. Goodfellow *et al.*, ―Generative adversarial nets,‖ in *Proceedings of advances in neural information processing systems*, 2014.

[17]. D. P. Kingma and M. Welling, ―Auto-encoding variational bayes,‖ in *Proceedings of international conference on learning representations*, 2013.

[18]. R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, ―An introduction to digital face manipulation,‖ in *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks*, Springer International Publishing Cham, 2022, pp. 3–26.

[19]. V. V. V. N. S. Vamsi *et al.*, ―Deepfake detection in digital media forensics,‖ *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 74–79, Jun. 2022, doi: 10.1016/j.gltp.2022.04.017.

[20]. R. Cellan-Jones, ―Deepfake videos double in nine months,‖ *BBC Tech News*. 2019.

[21]. D. Citron, ―How DeepFake Undermines Truth and threaten democracy,‖ in *TEDSummit*, TED official website, 2019. [Online].

[22]. Available: https://www.ted.com/talks/danielle_citron_how_deepfakes_undermine_truth_and_threaten_democracy

[23]. P. Korshunov and S. Marcel, ―DeepFakes: a New Threat to Face Recognition? Assessment and Detection,‖ *ArXiv*, vol. abs/1812.0, 2018.

[24]. B. B. C. Bitesize, ―Deepfakes: what are they and why would i make one,‖ *BBC Bitesize Articles*. 2019. [Online]. Available: https://www.bbc.co.uk/bitesize/articles/zfkwcqt

[25]. J. Kietzmann, L. W. Lee, I. P. McCarthy, and T. C. Kietzmann, ―Deepfakes: Trick or treat?,‖ *Bus. Horiz.*, vol. 63, no. 2, pp. 135–146, Mar. 2020, doi: 10.1016/j.bushor.2019.11.006.

[26]. S. Kolagati, T. Priyadharshini, and V. Mary Anita Rajam, ―Exposing deepfakes using a deep multilayer perceptron – convolutional neural network model,‖ *Int. J. Inf. Manag. Data Insights*, vol. 2, no. 1, p. 100054, Apr. 2022, doi: 10.1016/j.jjimei.2021.100054.

[27]. G. Wang, Q. Jiang, X. Jin, and X. Cui, ―FFR_FD: Effective and fast detection of DeepFakes via feature point defects,‖ *Inf. Sci. (Ny).*, vol. 596, pp. 472–488, Jun. 2022, Doi: 10.1016/j.ins.2022.03.026.

[28]. P. Korus, ―Digital image integrity – a survey of protection and verification techniques,‖ *Digit. Signal Process.*, vol. 71, pp. 1– 26, Dec. 2017, doi: 10.1016/j.dsp.2017.08.009.

References

[1]. H. Al-Najjar, S. Alharthi, and P. K. Atrey, —Secure image sharing method over unsecured channels,‖ *Multimed. Tools Appl.*, vol. 75, no. 4, pp. 2249–2274, Feb. 2016, doi: 10.1007/s11042-014-2404-5.

[2]. M. E. Hodeish, L. Bukauskas, and V. T. Humbe, —A new efficient TKHC-based image sharing scheme over unsecured channel,‖ *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1246–1262, Apr. 2022, doi: 10.1016/j.jksuci.2019.08.004.

[3]. J. Faircloth, —Information Security,‖ in *Enterprise Applications Administration*, J. Faircloth, Ed., Boston: Elsevier, 2014, pp. 175–220. doi: 10.1016/B978-0-12-407773-7.00005-3.

[4]. M. Devipriya and M. Brindha, —Secure Image Cloud Storage Using Homomorphic Password Authentication with ECC Based Cryptosystem Devipriya,‖ *Adv. Syst. Sci. Appl.*, vol. 22, no. 1, pp. 92–116, 2022.

[5]. V. L. Schultz, V. V Kulba, O. A. Zaikin, A. B. Shelkov, and I. V Chernov, —Regional Security: Analysis of the Emergency Management Effectiveness Based on the Scenario Approach,‖ *Adv. Syst. Sci. Appl.*, vol. 17, no. 1, pp. 9–24, 2017.

[6]. H. Kaur and S. R, —VLSI Implementation of Lightweight Cryptography Algorithm,‖ *Adv. Syst. Sci. Appl.*, vol. 16, no. 1, pp. 95–101, 2016.

[7]. R. Thabit and B. E. Khoo, —Robust Reversible Watermarking Application for Fingerprint Image Security,‖ *Adv. Syst. Sci. Appl.*, vol. 22, no. 1, pp. 117–129, 2022.

[8]. R. Thabit, —Improved steganography techniques for different types of secret data,‖ *Adv. Syst. Sci. Appl.*, vol. 19, no. 3, 2019.

[9]. S. Kloppenburg and I. van der Ploeg, —Securing Identities: Biometric Technologies and the Enactment of Human Bodily Differences,‖ *Sci. Cult. (Lond).*, vol. 29, no. 1, pp. 57–76, Jan. 2020, doi: 10.1080/09505431.2018.1519534.

[10]. J. Galbally, S. Marcel, and J. Fierrez, —Biometric Antispoofing Methods: A Survey in Face Recognition,‖ *IEEE Access*, vol. 2, pp. 1530–1552, 2014, doi: 10.1109/ACCESS.2014.2381273.

[11]. ISO/IEC JTC1 SC37 Biometrics, —Information Technology-Biometric Presentation Attack Detection-Part 3: Testing and Reporting.,‖ *Int. Organ. Stand.*, 2017.

[12]. S. Marcel, J. Fierrez, and N. Evans, *Handbook of Biometric Anti-Spoofing - Presentation Attack Detection and Vulnerability Assessment*. Springer Nature Singapore, 2023.

[13]. R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, —Deepfakes and beyond: A Survey of face manipulation and fake detection,‖ *Inf. Fusion*, vol. 64, no. July, pp. 131–148, Dec. 2020, doi: 10.1016/j.inffus.2020.06.014.

techniques capable of differentiating between authentic and manipulated face images. Additionally, further research is needed to enhance the capabilities of FIA schemes, particularly in the context of DeepFakes and other advanced manipulation methods. This may involve exploring the potential of deep learning-based techniques and the development of more comprehensive and versatile manipulation detection algorithms. Moreover, the paper suggests investigating the distinctions and interrelations among image tampering, manipulation, and forgery, and the development of tampering detection techniques with unique strengths and applications. Finally, the paper emphasizes the importance of addressing unanswered research questions and forecasting future directions to advance digital face manipulation generation and detection. By addressing these avenues for future research, the field of FIA and manipulation detection can continue to evolve and effectively mitigate the security risks associated with digital face manipulation methods.

Deep learning-based schemes require large datasets for training, and optimal detection performance is achieved when the input image closely matches the training set [120], [128].	**It can be applied to any type of image, and training is not required.**
The majority of methods employ labor- and time-intensive supervised networks for detection [129]–[131].	**The system is superior in terms of time and effort savings as it operates in a fully automated manner.**
Results of false detections have been documented, particularly in cases where the input image deviates from the training dataset [120], [128], [132], [133]. For example, the maximum accuracy values are 84.7%, 99.3%, 87.5%, and 81.6%, respectively [120], [132].	**There are no false detections, and the system exhibits high detection accuracy.**

5. Conclusion

This paper has highlighted the increasing significance of FIMD and FIA techniques in the context of digital data security. The emergence of digital face manipulation methods, including the notable development of "DeepFakes," has underscored the need for robust authentication and manipulation detection mechanisms. The limitations of current forensic software and the challenges posed by the rapid advancements in facial image manipulation techniques have been addressed. Furthermore, the paper has provided an overview of recent trends in FIA techniques and their limitations, emphasizing the need for continued development and innovation in this field. For future work, it is imperative to address the limitations of existing forensic software and to develop more effective manipulation detection

TABLE 5. Summary of some limitations in watermarking based FIMD and FIA.

Ref.	Scheme	Limitations
[107], [108]	FIMD	The original face region cannot be recovered after the manipulation-revealing process, although the watermarking-based technique can detect various manipulations.
[109]–[111]	FIA	The primary drawback of this scheme is considered to be its strict embedding capacity. Large face regions in the image can result in generating a significant number of bits for tamper detection, localization, and face region recovery, which require additional embedding space. If there is insufficient embedding capacity, the FIA scheme can not effectively protect the face image.

I. Comparison between deep-learning-based and watermarking-based FIMD techniques

Watermark-based FIMD and FIA techniques have proven efficiency and advantage over deep learning-based FIMD techniques. Table 6 presents a comparison of deep learning-based and watermarking-based techniques.

TABLE 6. Comparison between deep-learning based and watermarking-based FIMD techniques

Deep-learning-based techniques	Watermarking-based techniques
Most of these techniques are limited to detecting a single kind of manipulation because they rely on techniques for fabricating or manipulating images for implementation [50], [113]–[116], [127] .	Capable of identifying various forms of manipulation and do not reliant on knowledge of the processes involved in producing manipulated or fake images.

TABLE 4. Summary of some limitations in deep-learning based FIMD.

Ref.	Detected features	Limitations
[50], [112], [113]	Face Asymmetries	1- Small details often go unnoticed. 2- Accurate detection requires high-resolution images, which are challenging to find on social media networks. 3- It involves a high level of computational complexity. 4- It can identify a specific type of manipulation.
[114]	Landmark locations	1- Relying on landmark locations in face photos as a discriminative feature is not always dependable. 2- When sharing images over networks, having the front face view is not always necessary. 3- It can identify a specific type of manipulation.
[115], [116]	Color features	1- A substantial training set is necessary. 2- It involves a high level of computational intricacy. 3- Abundant computational capacity is required. 4- It can identify a specific type of manipulation.
[117]	Spatial artifacts	1- A sizable training set is necessary. 2- Training takes a considerable amount of time. 3- The computation techniques involved are challenging. 4- The precision and capability of manipulation detection are limited.
[104], [118]–[123]	Different features as inputs to CNN	1- A substantial training set is necessary. 2- Training takes a considerable amount of time. 3- Involves complex computation techniques. 4- The precision and capability of manipulation detection are limited.
[95]	Spectral distribution	1- Errors may occur when utilizing the energy spectral distribution. 2- Training takes a considerable amount of time. 3- Involves a convoluted calculation method. 4- It can identify a specific type of manipulation.
[124]–[126]	Spectral artifacts fitting parameters	1- Intricate computation techniques. 2- It took a significant amount of time to find the fitting parameters during post- processing. 3- Is limited to identifying GAN-based and certain retouching manipulations.

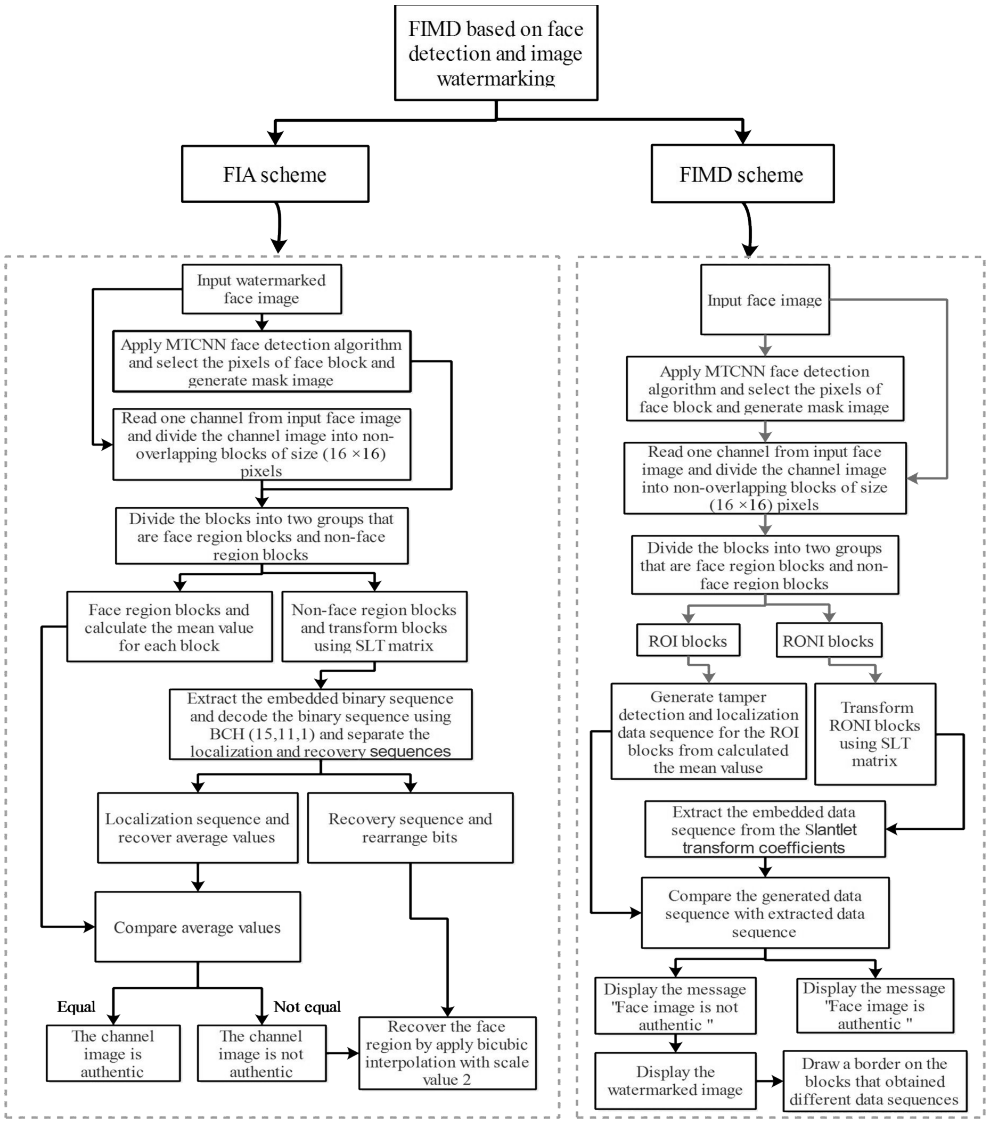


Fig. 4: Comparison of FIMD and FIA schemes on the receiver side of a single channel [108], [110].

tampered blocks in the face region and restores the original face region when manipulations are present [109]–[111]. A comparison of the sender and recipient sides of the FIMD and FIA schemes is provided in Figures 3 and 4. The original face data cannot be recovered by the FIMD scheme, which is intended to identify and detect tampering. However, the FIA scheme is able to recover the original face region in addition to detecting and locating tampering.

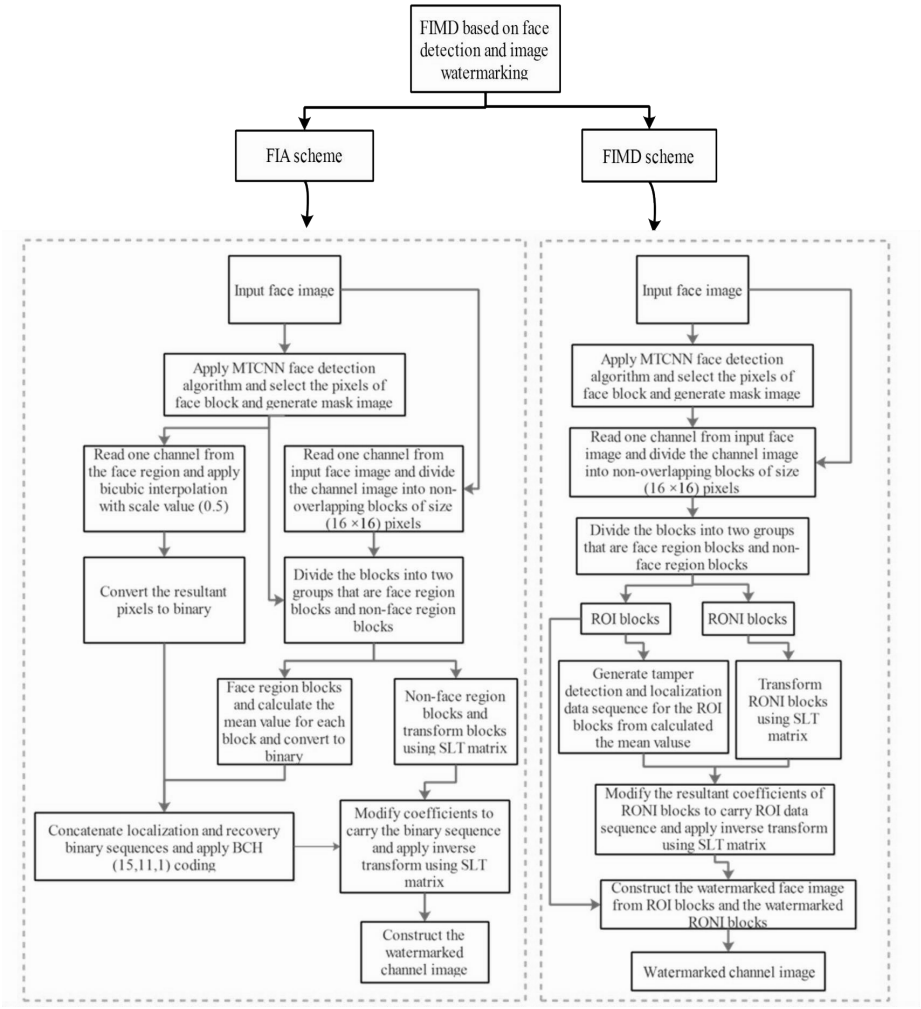


Fig. 3: Comparison of FIMD and FIA schemes on the sender side of a single channel [108], [110].

D. FIMD face detection and image watermarking

Another FIMD method that employs facial identification and picture watermarking algorithms was just released. Following the identification of the face region, the picture blocks are divided into two categories: those connected to the face region and those that are not. This categorization is achieved by creating a binary mask picture. Information is extracted from blocks in the face region and inserted into blocks outside of it using Slantlet-based picture watermarking, as the modification procedure demonstrates [107], [108]. Table 5 highlights some of FIMD's restrictions based on watermarking.

A Multi-Task Cascaded Neural Network (MTCNN) is used in a new FIA scheme to identify and choose the face region, after which output adjustments are made. Following the identification of the face region, a binary mask image is produced, which makes it easier to divide image blocks into two categories: face blocks and non-face blocks. The mean of every block from the face region is used to derive the tamper localization data. The recovery data from the face region is then produced using the Bicubic Interpolation (BI) algorithm. The system embeds the generated binary sequence into the non-face blocks using a content-based embedding algorithm that can embed binary bits in the High-Low (HL) and Low-High (LH) sub-bands of the Slantlet transform (SLT) coefficients of the block. When manipulations are present, the system. The system embeds the generated binary sequence into the non-face blocks using a content-based embedding algorithm that can embed binary bits in the High-Low (HL) and Low-High (LH) sub-bands of the Slantlet transform (SLT) coefficients of the block. The system consistently detects

A. FIMD based on texture analysis

The study focuses on assessing the susceptibility of face recognition systems to morphing attacks and proposes a novel technique for detecting such attacks. The suggested approach combines multiple Multi-scale Block Local Binary Patterns (MB-LBP). The study emphasizes that training and evaluating face morphing attack detection algorithms depend on robust morphing detection algorithms and the utilization of diverse databases [102].

B. FIMD based on digital forensics

The study investigates the impact of face retouching on face recognition and proposes a detection method based on Photo-Response Non-Uniformity (PRNU) for retouched face photos. The research assesses biometric performance both before and after retouching, coupled with a qualitative evaluation of beautification apps. The findings suggest that facial retouching only marginally decreases comparison scores. However, accurate identification of retouched face photos is crucial for upholding laws against Photoshop manipulation [103].

C. FIMD deep learning or artificial intelligence

According to its definition, Deepfakes is a deep learning-based method that replaces a person's face with another person's face to produce fake videos. The method creates realistic images and videos using GANs that can be exploited for financial fraud, fake news, and other malicious activities [14], [31], [40], [66], [104], [105]. Table 4 summarizes some limitations of FIMD based on deep learning.

4. FIMD techniques

The face recognition system (FRS) is a subset of biometric security software that also includes voice recognition, fingerprint recognition, and iris recognition [55], [78], [79]. The FIM process affects image quality, which can influence acceptance or rejection decisions in FRS [80]–[85]. When a fake image is accepted in FRS, it poses a threat to the entire security system and is viewed as a significant challenge to privacy control. On the other hand, intentional FIM attacks can lead to a variety of issues, including identity theft [86]–[89], political conflict as a result of fake news [66], [67], [90]–[92], financial fraud, and others.

Since the number of harmful FIM applications is rapidly increasing, the researchers have directed their efforts towards presenting manipulation detection techniques that can distinguish fake face images from authentic images [14], [66], [74], [93]–[95]. In recent years, several FIMD algorithms have been presented, which can be broadly classified into two types: differential-based algorithms and no-reference-based algorithms [96]–[102]. The differential detection algorithms require two images: the original image (the reference image) and the test image. These algorithms demonstrated their effectiveness in detecting manipulation; however, in traditional image forensics, there is only one video or image. Several studies have been directed toward detecting tampering when the trusted reference image is not available [40], [102]–[106].

Currently, four types of FIMD schemes are available:

The FIM can be used for a variety of purposes, both harmless and harmful [14], [66], [13]. Table 3 provides a brief description of the applications in which the FIM can be utilized.

TABLE 3. Description of FIM Example Applications

No.	FIM Application	Description
1	Spread Fake News	Images and videos are manipulated and edited to spread fake news on internet websites, magazines, and social media [67].
2	Digital Communication	Images are used in a variety of communication applications, including online registration and social media friendship [68].
3	Security Applications	Services include identity verification, online access control, identification, and authentication. [69].
4	Entertainment	The FIM is used in entertainment to create videos, advertisements, and other content. Some companies specialize in creating amusing cartoons from still images using lip-syncing technology [70]–[72].
5	Film Industry	Deepfakes have been used in the film industry for a variety of purposes, including changing the identity of the person in the recorded video and lowering costs by generating characters with AI tools [32], [73]–[75].
6	E-Commerce	Retailers have used deepfake technology to create tools that allow customers to swap their faces with digital models in virtual changerooms, potentially increasing online sales. Furthermore, ideal fake models generated using AI technology are used for advertisements at a significantly lower cost than real models [76].
7	E-Learning	Deepfake technology has the potential to improve children's education in a variety of ways, including swapping the teacher's face with their parents to help them coalesce and reinforce learning [66]. On the other hand, instructors can use the text and previous videos to create new video courses [77].

in Table 2, which provides a brief explanation of the common manipulation types. Fig. 1 presents the classification of the FIM.

TABLE 2. The common types of FIM techniques

No.	Manipulation type	Explanation
1	Entire face synthesis	Utilizes Generative Adversarial Networks (GANs) to generate virtual human faces that closely resemble real human faces. The generated faces can be used in various applications such as email, games, teleconferences, chat rooms, and various other contexts [16], [17].
2	Identity swap	This manipulation process involves replacing the person's face image with another face image, which can be used in a variety of applications, including the film industry, financial fraud, and video fabrication. The identity swap manipulation methods are divided into two categories: a) classical computer graphics-based techniques [51] and b) deep learning-based techniques [52].
3	Morphing	In this type of manipulation, a single morphed image is created by combining images from two or more people using specific software tools such as Face Morph‖, Face Swap Online‖, and the —morphed thing‖ [53]. The process of face morphing is focused on creating fake samples for photos, not for video. Furthermore, the front view of the face is usually considered in the process of manipulation [54]–[58].
4	Attribute manipulation	This manipulation process, also known as —face editing‖ or —face retouching‖ is used to change images based on attributes such as the color of individuals' hair or skin, their age, gender, the addition of a beard or glasses, and so on [59]–[61]. In most cases, this type of manipulation is considered an unintentional attack and is used to improve image quality. One example of this type of manipulation is the popular 'Snapchat' mobile application [62].
5	Expression Swap	The process of replacing an individual's facial expressions in a video clip with those of another individual [59], [60].
6	Audio-to-Video and Text-to-Video Swap	This type of manipulation is based on artificial intelligence (AI) techniques, and certain aspects must be considered, such as sounds, 3D face pose, expression, and scene light [63]–[65].

Due to the rapid advancements in the field of FIA, this paper is presented to review recent trends in FIA technologies and their limitations. Additionally, it aims to compare technologies based on deep learning and watermarking.

3. Types of FIM techniques

As previously stated, the process of FIM has become a prevalent topic in the last few years, especially after the term "DeepFakes" was recently distributed and highlighted by state-of-the-art research [13], [18]–[20], [22], [26], [47]–[50]. Several FIM techniques exist, and new manipulation methods are constantly being introduced through improved applications. Nevertheless, the majority of research has focused on the manipulation methods depicted

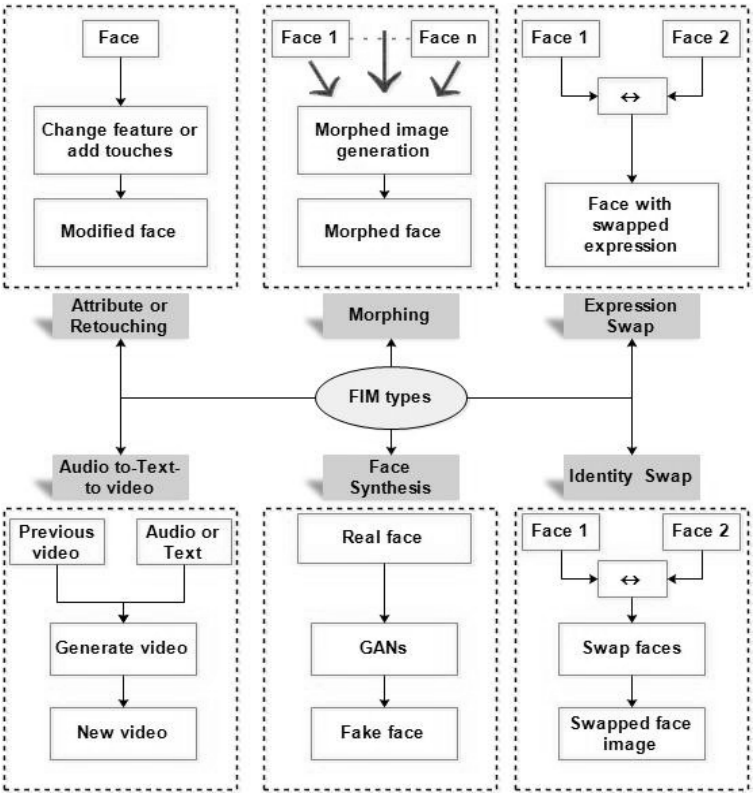


Fig. 1: Common types of FIM techniques [111].

Ref.	Year	Contributions
[39]	2021	The study investigates various state-of-the-art neural networks, such as MesoNet, ResNet-50, VGG-19, and Xception Net, with a specific emphasis on addressing complex issues posed by deepfakes generated through neural networks. To identify the most effective outcomes, the paper includes comparisons among the aforementioned methods.
[40]	2021	- Investigated the most recent techniques for creating and identifying deepfake photos. - Focused on summarizing and scrutinizing the architectures of methods for both generating and detecting deepfakes. - Presented future directions aimed at enhancing the architecture of deepfake models.
[41]	2021	- Aimed to present the latest findings in the identification and production of deepfake videos. - Evaluated the generalization and robustness of deepfake video generation and detection models. - Outlined the existing benchmarks for deepfake video creation.
[42], [43]	2021, 2022	A comprehensive analysis of deepfake detection using deep learning techniques: - Conventional neural network (CNNs)-based techniques. - Generative adversarial networks (GANs)-based techniques. - Recurrent Neural Network (RNN). - Long short-term memory (LSTM). - Autoencoder-based techniques.
[44]–[46]	2021, 2022	- Explored the latest techniques in the creation and detection of deepfakes. - Presented current datasets related to deepfake technology. - Examined prevalent issues and patterns associated with the production and identification of deepfakes.
[33]	2023	- Categorizes and discusses the latest research on face manipulation detection and generation. - Covering over 160 studies, it provides a comprehensive discussion of the subject. - The primary focus of the paper is on the creation and detection of deepfake content using deep learning. - Identifying obstacles, addressing unanswered research questions, and forecasting future directions contribute to the advancement of digital face manipulation generation and detection.

Table 1. Review papers summary

Ref.	Year	Contributions
[24]	2020	- Explained what a deepfake is and provided a rundown of the core technologies. - Categorized deepfake and IoT techniques and examined the opportunities and risks associated with each group. - A framework was established for successfully addressing the risks of deepfakes.
[13], [34]	2020	- Conducted a comprehensive analysis of contemporary face manipulation techniques, encompassing deepfakes and IoT, and explored methods for detecting them. - Classified these techniques into four major categories based on their prevalent use in facial manipulation. - Examined public`cly accessible datasets and essential benchmarks for detecting manipulated faces and assessing face manipulation techniques.
[35]	2020	- Conducted a comprehensive analysis of prevalent image forgery techniques. - Provided an overview of publicly available data sources for research on the identification of image manipulation. - Emphasized a primary focus on deep learning-based methods for detecting image manipulation.
[36]	2020	- Conducted an in-depth investigation to detect alterations in both photos and videos. - Emphasized the newly identified deepfake phenomenon from a forensic analyst's perspective. - Outlined the limitations of the latest forensic software, addressed pressing issues, examined imminent challenges, and proposed avenues for new lines of inquiry.
[37]	2020	- Reviewed the recent advancements and applications of semantic manipulation and face synthesis based on deep learning. - Discussed future perspectives aimed at further enhancing face perception through continued development and innovation.
[38]	2020	- Outlined the distinctions and interrelations among image tampering, image manipulation, and image forgery. - Provided justifications for various tampering detection techniques, highlighting their unique strengths and applications. - Investigated prevalent benchmark datasets commonly used in the assessment of tampering detection methods.

create fake digital material. To substitute the faces of celebrities in pornographic movies, a machine learning system known as DeepFakes was created in 2017 [25]. DeepFakes' capabilities have been utilized for a number of detrimental objectives, such as financial fraud, the dissemination of false information, and the production of sexual content [26]. Researchers have concentrated on creating face modification detection systems in an attempt to improve media forensics generally [27]–[32].The field of manipulated facial image generation is still in its early stages of development. Many papers provide a thorough examination of the methods used to create images of manipulated faces, with a particular emphasis on deepfakes and emerging techniques for detecting fake images [33]. The next section presents a summary of the recently published review papers in the field of face image manipulation and its detection techniques. To date, no review paper focuses on the recent FIA schemes; therefore, this paper presents a review of recent trends in FIA techniques and their limitations. The rest of the paper is organized as follows: section II presents a summary of some review papers that are related to the FIM and FIMD schemes; section III presents the types of FIM techniques; section IV presents a review of face image manipulation detection (FIMD) techniques; section V presents a comparison between deep-learning based and watermarking based FIMD techniques; and section VI presents the conclusions and suggestions for future researches.

2. Literature Survey

The study in this section will present a summary of recently published review papers related to the field of FIM and FIMD, including references, year of publication, and initial contributions, as shown in Table 1.

1. Introduction

As technology develops, digital photos and IoT may now be shared easily for a variety of reasons [1]–[3]. These days, people who utilize technology find that storing their private information and images in the cloud is really convenient because they can access it anytime they need to. However, security and data integrity are the main issues that users are worried about [4], [5]. To ensure the security of digital photos, the Internet of Things, and shared data, a number of techniques and security frameworks have been put into place, such as watermarking, steganography, and cryptography [6]–[8 Face images, which are shared online for a variety of reasons, including social media, face recognition apps, celebrity and fame aspirations, and identity discrimination in biometric systems (i.e., security and access control), are among the most important types of digital images and IoT [9]. There are several other uses for face photographs [10]–[12]. Because of the speed at which technology is developing, digital face editing techniques, algorithms, and applications are now readily available [13]–[17]. Intentional attacks, which involve malicious intent during the manipulation process, and unintentional attacks, which involve harmless intentions like beautifying the face, adjusting lighting, adding humorous stickers, etc., are the two main categories into which techniques for manipulating facial images can be divided. The content of the face picture is changed (i.e., its features) as a result of both modification approaches [18]. Interest in FIM techniques and how they affect data security systems has increased in the data security community. A vibrant field of study has emerged as a result of the scholarly community's attention being drawn to the term "DeepFakes" in recent years [19]–[24]. Known as DeepFakes, deep learning techniques have been used to

المستخلص

يُبرز التطور السريع لإنترنت الأشياء وخوارزميات التلاعب بصور الوجوه (FIM)، بالإضافة إلى نمو تطبيقاتها سهلة الاستخدام، الحاجة المُلحة لأساليب كشف التلاعب. يجب أن توضح هذه التقنيات كيفية تعديل صور الوجوه والتحقق من صحتها. وقد انتبه المجتمع العلمي مؤخرًا إلى مصطلح "التزييف العميق" وطرق كشفه. انتبه أيضًا إلى أحدث الطرق لكشف تعديلات صور الوجوه القائمة على العلامات المائية. من المهم تذكر أن لكل طريقة من هذه الطرق عيوبها. تُقدم هذه الدراسة مقدمة موجزة عن أساليب كشف تعديل صور الوجوه، مُركزةً على كلٍّ من الأساليب القائمة على العلامات المائية والتعلم العميق. بعد ذلك، تُقدم الورقة أمثلةً تُوضح استخداماتها، مُشددةً على مُقارنة بين التعلم العميق والتعلم العميق. ثم تُقدم الورقة أمثلةً على كيفية استخدامها، مُركزةً على مُقارنة أساليب التعلم العميق والعلامات المائية للكشف عن التلاعب. وتضمنت خاتمة الورقة توصيات بحثية ورؤىً حول التطورات المُستقبلية المُحتملة في هذا المجال الدراسي المُثير للاهتمام. للباحثين المهتمين أو النشطين في هذا المجال الدراسي، تُعدّ المراجعة الواردة في هذه المقالة مرجعًا أساسيًا.

الكلمات المفتاحية: إنترنت الأشياء (IoT)، مصادقة صورة الوجه (FIA)، التزييف العميق. (Deepfakes).

Abstract

The quick development of IoT and facial image manipulation (FIM) algorithms, as well as the growth of their user-friendly applications, highlight the pressing need for manipulation detection methods. These techniques need to demonstrate how face photos have been altered and validate their legitimacy. The scientific community has recently taken notice of the phrase "DeepFakes" and methods for detecting them. Take note of the latest methods for identifying watermark-based face image modification as well. The important thing to remember is that every one of these methods has its own set of drawbacks. This study provides a brief introduction to face image modification detection methods, emphasizing both watermarking-based and deep learning-based methods. Afterwards, the paper offers instances that demonstrate their use, stressing a comparison between deep learning. The paper then goes on to give examples of how they might be used, with a focus on comparing deep learning and watermarking-based methods for detecting tampering. Research recommendations and insights into possible future advancements in this fascinating field of study are included in the paper's conclusion. For scholars actively involved in or interested in this particular field of study, the review in this article is regarded as an essential resource.

Keywords: IoT (Internet of Things), Face Image Authentication (FIA), DeepFakes.

Examining IoT-Enhanced for Current Developments in Face Image Authentication (FIA) Methods and Their Drawbacks / Review article

Marwa Jamal Hadi[1] and Emaan Ouudha Oraby[2]

1 Fine Arts Institutes for Girls, Baqubah-Diyala / Iraq

2 General Directorate of Education in Al-Qadissiyah Government, Ministry of Education / Iraq

e-mail1: marwajmal178@gmial.com

e-mail2: eman.iraq2017@gmail.com

دراسة تقنيات إنترنت الأشياء المحسنة للتطورات الحالية في أساليب مصادقة صورة الوجه (FIA) وعيوبها ـ دراسة مرجعية

مروة جمال هادي[1] و إيمان عودة عرابي[2]

1 معاهد الفنون الجميلة للبنات، بعقوبة \ ديالى – العراق

2 المديرية العامة للتربية في محافظة القادسية – وزارة التربية والتعليم \ العراق

[31]. S. Obermeyer *et al.*, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science*, vol. 366, no. 6464, pp. 447–453, 2019.

[32]. A. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A Survey on Bias and Fairness in Machine Learning," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–35, 2021.

[33]. M. Zech *et al.*, "Variable Generalization Performance of a Deep Learning Model to Detect Pneumonia in Chest Radiographs: A Cross-Sectional Study," *PLoS Med.*, vol. 15, no. 11, e1002683, 2018.

[34]. S. Irvin *et al.*, "CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison," in *Proc. AAAI*, vol. 33, no. 1, pp. 590–597, 2019.

[35]. D. B. Saria, "Individualized Sepsis Treatment Using Reinforcement Learning," *Nature Medicine*, vol. 25, pp. 1436–1445, 2019.

[36]. W. Xie *et al.*, "Multimodal Learning for Healthcare: A Review," *IEEE Trans. Artif. Intell.*, vol. 2, no. 4, pp. 446–460, 2021.

[37]. A. Joshi *et al.*, "Machine Learning for Global Health: Lessons from COVID-19 and Beyond," *Lancet Digit. Health*, vol. 3, no. 6, pp. e340–e341, 2021.

[38]. L. Morley *et al.*, "The Ethics of AI in Health Care: A Mapping Review," *J. Med. Internet Res.*, vol. 22, no. 8, e16228, 2020.

[39]. Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, 2019.

[40]. B. Sendak *et al.*, "A Path for Translation of Machine Learning Products into Healthcare Delivery," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–10, 2020.

[41]. S. Wiens *et al.*, "The Future of AI in Healthcare: A Review of the Regulatory Landscape," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 4, pp. 1022–1029, 2021.

[15]. N. Rieke *et al.*, "The future of digital health with federated learning," *NPJ Digital Medicine*, vol. 3, no. 119, pp. 1–7, Sept. 2020.

[16]. B. A. Yu, A. Beam, and I. S. Kohane, "Artificial intelligence in healthcare," *Nature Biomedical Engineering*, vol. 2, no. 10, pp. 719–731, Oct. 2018.

[17]. M. J. van der Schaar *et al.*, "How artificial intelligence is changing the role of the doctor," *BMJ*, vol. 363, p. k3797, Oct. 2018.

[18]. J. D. Kelleher, *Machine Learning: A Concise Introduction*, MIT Press, 2019.

[19]. A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[20]. J. Lee *et al.*, "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020.

[21]. M. Abadi *et al.*, "TensorFlow: A System for Large-Scale Machine Learning," in *Proc. 12th USENIX Conf. Oper. Syst. Design Implement.*, 2016, pp. 265–283.

[22]. P. Gope and B. H. Kim, "A Survey on Security and Privacy Issues in Modern Healthcare Systems: Attacks and Defenses," *IEEE Access*, vol. 7, pp. 88521–88599, 2019.

[23]. V. Gulshan *et al.*, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.

[24]. P. Rajpurkar *et al.*, "Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks," *Nature Medicine*, vol. 25, no. 1, pp. 65–69, 2019.

[25]. D. Libbrecht and W. S. Noble, "Machine Learning Applications in Genetics and Genomics," *Nature Reviews Genetics*, vol. 16, pp. 321–332, 2015.

[26]. H. J. Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*, Basic Books, 2019.

[27]. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?" Explaining the Predictions of Any Classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 1135–1144.

[28]. S. Holzinger *et al.*, "What Do We Need to Build Explainable AI Systems for the Medical Domain?" *arXiv preprint arXiv:1712.09923*, 2017.

[29]. D. R. Harder *et al.*, "Towards Clinically Applicable Explainable AI Models in Healthcare," *npj Digit. Med.*, vol. 4, no. 1, pp. 1–6, 2021.

[30]. E. Choi *et al.*, "RETAIN: An Interpretable Predictive Model for Healthcare Using Reverse Time Attention Mechanism," in *Proc. NIPS*, 2016, pp. 3504–3512.

References

[1]. A. Esteva *et al.*, “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, vol. 542, pp. 115–118, Feb. 2017.

[2]. V. Gulshan *et al.*, “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.

[3]. M. S. Hossain and G. Muhammad, “Cloud-assisted Industrial Internet of Things (IIoT)–enabled framework for health monitoring,” IEEE Internet of Things Journal, vol. 6, no. 5, pp. 8085–8093, Oct. 2019.

[4]. A. Yazdinejad *et al.*, “Real-time remote health monitoring via intelligent edge computing: A systematic review,” IEEE Access, vol. 9, pp. 57005–57021, 2021.

[5]. J. D. Power *et al.*, “Ethical considerations in artificial intelligence applications in clinical diagnosis,” IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 5, pp. 1335–1343, May 2020.

[6]. E. H. Shortliffe and B. G. Buchanan, "A model of inexact reasoning in medicine," *Mathematical Biosciences*, vol. 23, no. 3–4, pp. 351–379, 1975.

[7]. [K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4–5, pp. 198–211, Jun. 2007.

[8]. A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[9]. M. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 11, pp. 4793–4813, Nov. 2021.

[10]. J. Lee *et al.*, "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, Feb. 2020.

[11]. A. Alsentzer *et al.*, "Publicly Available Clinical BERT Embeddings," in *Proc. 2nd Clinical NLP Workshop*, Florence, Italy, Aug. 2019, pp. 72–78.

[12]. H. Rajkomar *et al.*, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, Apr. 2019.

[13]. M. Chen *et al.*, "Big Data Analytics for Personalized Healthcare," *IEEE Transactions on Emerging Topics in Computing*, vol. 2, no. 3, pp. 166–173, Sept. 2014.

[14]. Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, pp. 1–19, Feb. 2019.

healthcare professionals are able to provide more accurate and timely diagnostic results. With growing potential in edge computing and real-time data interpretation, AI systems can now assist with clinical decision-making at point-of-care to improve patient management. Nevertheless, despite these tremendous developments, there are several obstacles currently preventing the extensive implementation of AI in health-related diagnostics. However data quality and availability are major concerns and AI requires significant amount of high quality labeled data to be trained well. Moreover, the black-box property associated with the majority of deep learning models, limiting interpretability, may prevent clinicians from trusting and accepting AI-based decisions. There are also lingering ethical considerations around things like algorithmic bias and patient privacy, as AI models may inadvertently reinforce inequities in health outcomes. Despite the promise, to fully harness the power of AI in healthcare diagnostics, future research must focus on enhancing model interpretability, fairness, and adoption mechanisms that can seamlessly package AI into clinical workflows. Here we advocate for a multidisciplinary, collaborative framework that brings together clinicians, data scientists, ethicists, and policymakers to ensure that the development of AI systems is ethical, complies with regulations, and is trustworthy in real-world clinical settings. At the end of the day, the eventual fate of AI in healthcare diagnostics will come down to building reliable, democratic and open models that don't just improve diagnostic accuracy, but also advance health equity and improve patient outcomes all over the world.

9. Future Research Directions

As the integration of Artificial Intelligence (AI) in healthcare diagnostics continues to evolve, several key areas warrant focused research to ensure effective, ethical, and sustainable implementation. Future investigations should aim to address current limitations, improve system performance, and align AI tools with clinical needs and societal values. Key research directions include:

a) Develop generalizable models for reliable performance across populations and settings.
b) Create federated learning frameworks for collaborative training without data centralization.
c) Enhance explainability and clinician-AI collaboration.
d) Design user-friendly interfaces and decision pathways.
e) Integrate AI education into medical training for ease of adoption.
f) Evolve regulatory frameworks for continuous validation and post-deployment monitoring.
g) Prioritize cross-disciplinary studies for effective and socially responsible AI tools.

10. Conclusion

Artificial Intelligence (AI) has become a game changer for healthcare diagnostics with dramatic enhancement in accuracy, speed and accessibility. Through the wide-spread application of AI techniques such as deep learning, machine learning, and natural language processing (NLP) in various diagnostic realms, including medical imaging, genomics and biosignal analysis,

Real-time and edge AI implementation is an increasing trend, due to the evolution in IoMT. However, balancing the trade-offs between latency, accuracy and power consumption requires careful engineering. Moreover, the sparse inclusion of low resource languages and regions in the generation of language data sets also reflects on a worldwide equity issue [37].

Although XAI and fairness have increased in relevance, concrete methodologies for their application in the clinical practice are evolving. Multidisciplinary models with clinicians and data scientists and ethicists can help bridge a gap between what AI can do and what is needed [38]. So we can see what are the key trends and gaps and future directions for AI diagnostics research from Table 5.

Table 5: Emerging Trends and Unaddressed Gaps in AI Diagnostics Research

Category	Current Trends	Identified Gaps / Needs
Model Usage	**CNNs widely used for imaging; transformers gaining momentum in text and multimodal tasks [36]**	**Holistic multimodal integration (imaging + biosignals + text) is underexplored**
Deployment Platforms	**Increasing focus on real-time AI and edge computing through IoMT [36]**	**Engineering trade-offs: latency, accuracy, and power constraints require optimization**
Equity in Datasets	**Research mostly focuses on high-resource regions and major global languages [37]**	**Underrepresentation of low-resource languages/regions limits model generalizability**
Explainability & Fairness (XAI)	**Growing interest in fairness and explainability tools [38]**	**Lack of mature, deployable strategies in clinical practice; need for clinician engagement**
Collaboration & Integration	**Cross-disciplinary collaboration is recognized as important [38]**	**More structured frameworks involving clinicians, ethicists, and engineers are needed**

Deployment into the real world must confront interoperability, user experience, and regulatory approval problems, which are rarely touched upon in technical papers [35]. Table 4 summarizes the main methodological and practical challenges of current AI diagnostics studies.

Table 4: Limitations of Current Methodologies in AI Diagnostics and Proposed Solutions

Limitation Category	Description	Proposed Solution / Need
Data Limitations	**Most studies use retrospective, curated datasets that lack clinical diversity [33]**	**Conduct prospective, multi-center studies that reflect real-world scenarios**
Evaluation Inconsistency	**Lack of standardized benchmarks and inconsistent metrics across studies [34]**	**Develop standardized datasets, evaluation protocols, and public challenge frameworks**
Clinical Integration Gap	**Research often lacks consideration for real-world implementation (interoperability, UX, regulation) [35]**	**Emphasize deployment feasibility, clinician involvement, and regulatory compliance**
Reproducibility Issues	**Model comparisons are often non-reproducible due to diverse experimental setups**	**Require transparent reporting, shared code, and common testing standards**
Academic Silos	**Dominance of academic-led research with limited collaboration with hospitals and industry**	**Foster academia-clinic-industry partnerships to bridge research and real deployment**

8. Survey Insights and Research Gaps

This review indicates leading trends and growing voids in AI diagnostics research. CNNs are widely used in imaging tasks, and recently transformer-based architectures have gained popularity in both textual and multimodal domains. Yet, the holistic integration of multimodal data (including imaging, signals, clinical text data) lacks clear exploration [36].

[30]. Moreover, the scarce availabilities of large, annotated datasets prevent the training of cumbersome models, in particular for rare diseases categories.

Another issue is the interpretability of the model. Practitioners are typically wary of opaque algorithms, especially without detailed transparency on critical diagnostic calls. Explainable AI (XAI) frameworks are expected to mitigate this challenge by introducing model decisions in visual or rule-based forms, but their usage has been so far quite confined [31].

Algorithmic bias carries ethical and legal hazards. For example, if the training data is not diverse, models could perform poorly for particular demographic groups, which could subsequently worsen health disparities. The issue of fairness in AI would need to be carefully considered to include representative datasets and bias correction in decision making algorithms [32].

7. Limitations of Existing Research

Methods and practical constraints pose challenges in the AI diagnosis research. Most of the studies are retrospective, based on selected datasets, and do not represent clinical heterogeneity. Back in China, multi-center prospective trials will be required to prove AI systems are reliable under different operational circumstances [33].

The lack of common benchmarks is also a limitation in terms of the performance assessment. Comparison with other models is difficult since different datasets, metrics and evaluation procedures are used. Adoption of standardised datasets and challenge structures would facilitate this, enabling fair comparisons and progress to be made [34].

In addition, the majority of the research is carried out in the university environment and little attention is paid to the integration into the clinic.

Table 3: Key Differences Between AI-Driven and Conventional Diagnostic Approaches

Criteria	AI-Based Diagnostic Systems	Traditional Diagnostic Methods
Accuracy	Often equal to or higher than experts in image/text-based diagnosis [27]	Generally high, but may vary with human factors
Specificity & Sensitivity	High, especially in imaging and biosignal interpretation	Dependent on practitioner skill and standardized protocols
Time Efficiency	Real-time or near-real-time; scalable across large datasets	Time-consuming; manual processes
Interpretability	Often limited ("black box" models like deep learning)	High (transparent logic in rule-based systems)
Adaptability	Continuously improves with data; supports personalization	Rigid protocols; less adaptive to individual variation
Scalability	Easily deployed across cloud, edge, and IoMT devices	Limited to facility-based, manual workflows
Transparency	Requires explainable AI tools (e.g., SHAP, LIME)	Inherent transparency in logic-based systems
Regulatory & Validation	Requires ongoing validation and regulatory adaptation	Built upon well-established clinical guidelines
Hybrid Model Potential	High: Combines rule-based logic with learning capabilities for improved performance [28]	Low: Limited capacity to integrate learning mechanisms
Clinician Involvement	Requires trust-building, human-in-the-loop systems [29]	High involvement; clinicians are decision-makers

6. Challenges and Issues

There are many challenges associated with the use of AI i:n diagnostics. First is heterogeneity of data—health data differs across institutions, patient populations and modalities, making the generalizability of AI models challenging

Overview of main comparative findings for selected studies Table Key comparative insights from selected studies, AI out-performs in speed and scale and adds more value by not missing or improving on diagnostic performance [27].

Even though AI has exhibited its performance advantages, interpretability remains a notable aspect where conventional techniques still have advantages. There are transparent logical trails in the rule based expert systems, while deep learning models are “black-box”. Rule-based reasoning and data-driven learning hybrid models are becoming popular as a middle solution [28].

In addition, unlike established diagnostics based on sound validated protocols, AI would need to be revalidated constantly using new data to remain valid. The comparison highlights the importance of appropriately training AI into clinical workflow with regulation and clinician involvement in mind [29]. Table 3 summarize comparison between AI based diagnostic systems and conventional diagnostic systems.

4. Applications of AI in Healthcare Diagnostics

AI in health diagnostics span across range of fields such as medical imaging, biosignal processing, genomics and clinical decision support. In the field of imaging diagnostics, CNN based systems have reached the performance of a dermatologist for skin cancer detection and high sensitivity for diabetic retinopathy (from retinal scans.) [23] Automated pattern recognition techniques have also made significant contributions to pathology and radiology, greatly decreasing the manual burden and diagnostic variation.

Biosignal-based diagnosis using RNNs and attention mechanism to translate ECGs, EEGs, and other human physiologies in real-time, for early screening of cardiac arrhythmia and neuro-disease detection [24]. In genomics, AI is utilized for prediction of mutations, classification of variants, and identification of susceptibility to disease based on sequencing data [25].

AI-based Clinical decision and support systems Conventional Clinical Decision Support Systems (CDSS) that utilize AI technology, incorporate information from the patient, medical standards and immediate analysis to offer evidence-based advice and suggestions. Such systems both improve the diagnostic consistency and the patient outcome in challenging or equivocal clinical cases [26]. The broad usage of such tools indicates the desire of AI in enhancing human expertise across many diagnostic areas.

5. Comparative Analysis of AI Approaches

For the assessment of the effectiveness of AI systems with the conventional diagnostic strategies of pathology, there a few metrics accuracy, specificity, sensitivity and the time efficiency are taken into account. Table

further develops to more general deep learning (DL) methods. These methods are used to structured (e.g., images and sequential) and unstructured data (e.g., clinical text) for models such as Convolutional Neural Networks (CNNs). NLP models are crucial for deriving actionable knowledge from unstructured documents such as EHRs. Frameworks such as TensorFlow, PyTorch and Scikit-learn are enabling the development and deployment of these models. Last but not least, hardware support (such as those from GPUs and TPUs) and edge and cloud-based infrastructures allow real-time and decentralized diagnostics applications. Figure 2 Framework of AI Techniques and Infrastructure in Healthcare Diagnostics Figure 2 shows Framework of AI Techniques and Infrastructure in Healthcare Diagnostics

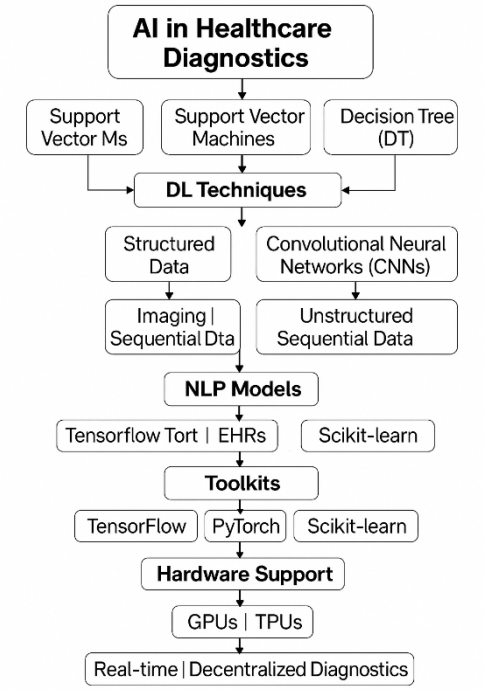


Figure 2: Framework of AI Techniques and Infrastructure in Healthcare Diagnostics

3. Techniques an d Tools Used in AI for Diagnostics

AI methods in healthcare diagnostics are diverse, as they are drawn from a variety of machine learning ML and deep learning (DL) algorithms developed for specific diagnostic tasks. Classical ML models, including Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF), have been shown to have good performance in structuring data classification tasks and risk stratification [18]. But the wave of success for DL models, especially CNNs and RNNs, has consequently changed the paradigm to process image-based and sequential data [19].

The Natural Language Processing (NLP) tools like BERT, BioBERT, and ClinicalBERT the diagnostic information can be extracted from the unstructured clinical texts, Electronic Health Record (EHR), etc., and it could be served as a serious help in decision-making process [20]. Furthermore, AI toolkits and platforms such as TensorFlow, PyTorch, and Scikit-learn establish interfaces for AI algorithm development and deployment.

The adoption of AI in healthcare diagnostics is also facilitated by highperformance hardware platforms like GPUs and TPUs [19] and edge computing infrastructure, which enables online processing on medical devices [21]. Advances in Internet of Medical Things (IoMT) brought diagnostic models closer to the patient bed by operating in decentralized settings, enabling the use of intelligent diagnostic tools in remote or poor resource areas [22].

This flowchart describes the technology ecosystem of AI in healthcare diagnostics, presenting how the different ML and DL methods are mainstreamed into diagnostic workflow. It starts with classic models including Support Vector Machines (SVMs) and Decision Trees (DTs) and

Table 2: Comparative Analysis of Key Literature on AI in Healthcare Diagnostics

Ref.	Study Focus	Techniques/ Models Used	Strengths	Limitations/Gaps
[6]	Rule-based expert systems in diagnostics	Knowledge-based systems	High interpretability; early integration in decision support	Poor scalability and adaptability to new data
[7]	Early machine learning in imaging diagnostics	Decision Trees, SVMs	Improved performance over rule-based systems	Heavy dependence on feature engineering
[8]	Deep learning in skin cancer classification	CNNs	Achieved dermatologist-level accuracy	Lack of explainability; risk of bias in datasets
[9]	Explainable AI (XAI) for medical imaging	Saliency maps, attention layers	Enhances interpretability of deep models	Still limited in clinical adoption due to complexity
[10]	NLP in biomedical literature and records	BERT, BioBERT	State-of-the-art clinical NLP performance	Requires large annotated corpora; domain-specific challenges
[11]	Clinical adaptation of NLP models	Clinical BERT	Better contextual understanding of clinical language	Limited generalization across institutions
[12]	Multi-modal data integration	EHRs, Imaging, Genomics	Supports personalized and holistic diagnostics	Data heterogeneity; alignment challenges
[13]	Big data analytics in healthcare	Cloud-based AI systems	Scalability and population-level insights	Privacy concerns; data siloing
[14]	Federated learning for privacy-preserving AI	Federated deep learning	Enables model training without central data sharing	Complex synchronization; performance variation
[15]	Edge computing in digital health	Edge AI architectures	Real-time inference and reduced latency	Hardware limitations; regulatory uncertainties
[16]	Ethical concerns and validation gaps	Conceptual frameworks	Highlights societal and regulatory implications	Underrepresentation in technical studies
[17]	AI's impact on clinical roles and collaboration	Review and analysis	Emphasizes interdisciplinary needs and clinician trust factors	Lack of concrete implementation strategies

these issues, approaches, such as synthetic data generation and federated annotation pipelines are being investigated.

On the deployment side, there is also an emerging literature on how to deploy AI models at the point of care in an optimal way. Approaches to edge processing and federated learning make it possible to run AI-driven diagnostics directly on medical devices or local servers, minimizing latency and keeping patient information private [14]. Faster architectures have been proposed to address these limitations; however, those solutions add additional overhead in distributed training, real-time inference, and synchronization. Moreover, enforcing healthcare-specific policies—e.g., HIPAA or GDPR—within decentralized settings continue to be a key challenge [15].

Whilst advances have been made in technology, the existing literature remains far from reaching today's demands. Such ethical considerations — especially regarding algorithmic bias, equity, transparency, and patient autonomy — are also frequently overlooked in technical articles. In addition, there is a lack of prospective, clinical trials that have to support the use of AI models in everyday diagnostic pathways [16]. Interdisciplinary cooperation among computer engineers, physicians, law professionals, and ethicists is lacking but is key to translating AI advancements to clinically safe, ethically sustainable, and regulatory acceptable tools[17].

To summarize the main contribution, methodology and identified gaps over reviewed literature, Table 2 provides the comparison and focus on areas, techniques used, strengths, and limitations of some major studies to provide a synthesized view of the current state of AI-based healthcare diagnostic.

been shown to reach clinician-level performance in radiology, dermatology, and ophthalmology. CNNs showed better performance in classification of skin lesions [6] and pneumonia detection in chest X-rays [8] than general practioneers. Model interpretability, however, categorically referred to as the "black-box" problem, causes difficulties in clinical trust and acceptance. Recent studies attempt to combine explainable AI (XAI) such as [1] attention maps and layer-wise relevance propagation with the learning process to improve the transparency and diagnostic accuracy [9].

Simultaneously, natural language processing (NLP) has developed as a powerful tool in the analysis of unstructured clinical text, such as discharge summaries and radiology reports. Compound models such as: BERT, and BioBERT achieved competitive results in named entity recognition, relation extraction and clinical text classification tasks [10]. However, the medical domain has its own challenges in terms of considering specialized vocabulary, non-standard documentation style and insufficient annotated data. Domain adaptation, self-supervised learning and continual fine-tuning are a current focus of research towards enhancing NLP robustness across institutions and patient populations [11].

An other important line of research, multi-modal diagnostics, refers to a situation where AI extracts features from multiple sources, such as imaging, genomics, electronic health records (EHRs), and biosignals. This will open up the path to a more comprehensive and individualized diagnostic concept. But, integrating disparate data sources is difficult itself, such as format reconciliation, time mismatch and sparse data [12]. Besides, the lack of such genuinely labeled and demographically diverse datasets hampers the generalizability and fairness of diagnostic models [13]. To combat

contributions in algorithmic innovation and hardware design, identified medical domains where AI has shown the most promising outcomes, and surveyed the architectural models in which AI deployments are conducted in the cloud, edge, and hybrid settings. Through a thematic organization of the literature and highlighting the directions of the emerging trends and existing gaps and challenges, this paper should serve as a helpful read for researchers, developers, and policymakers aiming to advance the incorporation of AI technologies into diagnostic. Lastly, this work hopes to contribute to the connection of AI breakthroughs to the actual medical practice, creating a pathway for next-generation diagnostic system creation that is smart and quick, but also ethical, interpretable, and aligned with the clinical settings.

2. Literature Review

Intelligence artificial (AI) has made an incredible headway in the diagnoses in health from the power to find complex patterns in medical data. The early AI systems were largely rule-based expert systems where professional expertise was encapsulated in deterministic decision trees. Although being transparent and interpretable, these systems lacked generalizability in different diagnostic settings and performed suboptimally on noisy or unstructured data [6]. The rise of machine learning (ML), in particular supervised learning algorithms, including support vector machines, and decision trees facilitated data-based modeling. These approaches enabled greater accuracy and efficiency, however required significant hand-crafted features and domain specific pre-processing [7].

The emergence of deep learning technology, and in particular CNNs, revolutionized AI diagnostics -- and especially medical imaging. CNNs have

Although there has been great progress, there are still obstacles to overcome. The heterogeneous nature of data, lack of high-quality labeled datasets, model interpretability and ethical considerations regarding algorithmic bias and patient privacy are some of the most critical barriers that are preventing large-scale clinical deployment of AI-based clinical solutions [5]. And what's more, the embedding of AI into existing healthcare infrastructure requires strong interoperability standards, regulatory approvals and clinician trust factors which are commonly overlooked in strictly technical research.

As a rough comparison between artificial intelligence systems and conventional diagnostic systems, Table 1 shows the improvement in accuracy, speed, scalability, and flexibility, in addition to real-life examples Table 1 Differences between traditional diagnostic systems and artificial intelligence systems.

Table 1: Comparison of AI and Traditional Diagnostic Systems

Aspect	Traditional Systems	AI-based Systems	Example
Diagnostic Accuracy	Heuristic/Experience-Based	Data-Driven, High Precision	AI in skin cancer classification
Speed	Manual Analysis	Real-time Processing	Retinal screening in seconds
Scalability	Limited to Specialist Availability	Deployable on Cloud/ Edge	Telemedicine platforms
Interpretability	Transparent Rules	Often Black-Box Models	Explainable AI efforts

In summary, this survey aims to provide a structured and comprehensive overview of the current prospects of AI in healthcare diagnostic, especially in the real-time intelligent systems. Specifically, this paper analyzed the recent

addition, the emergence of edge computing, and Internet of Medical Things (IoMT) make it possible to integrate diagnostic capacity into a wearable device, a mobile platform, or a remote monitoring tool to provide the continuous patient evaluation with the least latency [3], [4].

Artificial Intelligence in Healthcare Infographic 1 shows the role of AI in data acquisition, automated detection, and real-time decision support in various aspects of medical diagnosis at different levels of sample collection and data analysis. It incorporates sources of data (imaging, biosignals), AI models (CNNs, NLP, etc.and, computing platforms (cloud, edge, IoMT), as well as the ultimate diagnostic results.

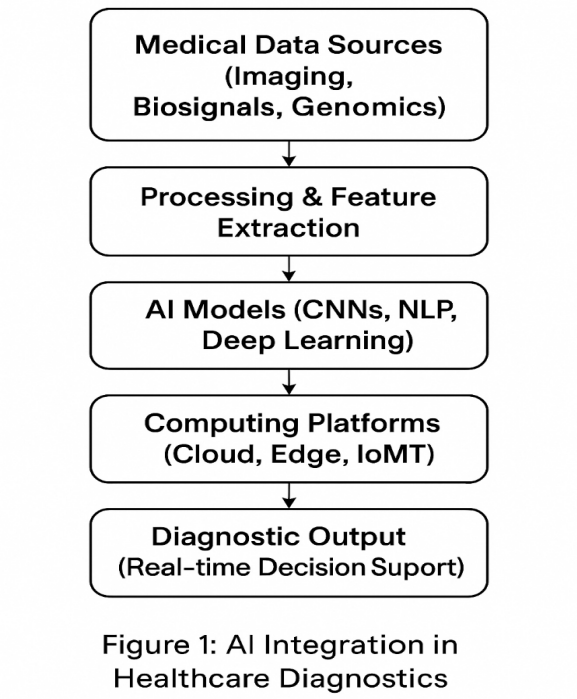


Figure 1: AI Integration in Healthcare Diagnostics

1. Introduction

Artificial Intelligence (AI) is quickly transforming healthcare, especially the medical diagnostic field where quick and precise decision making is crucial. As the size and complexity of clinical data, from medical images to genomic profiles, show no signs of slowing down the traditional diagnostic methods, often involving manual interpretation and rule-based heuristics, are falling behind the trend of modern healthcare. AI, infused with advancements in deep learning, machine learning and intelligent computing architectures, is now being harnessed in ways that have opened up new possibilities and levels of diagnostic accuracy, error reduction, and real-time clinical decision making.

In recent years, performance of AI models has rivaled and sometimes surpassed that of experienced clinicians in a wide range of diagnostic tasks. For example, the use of convolutional neural networks (CNNs) has reached dermatologist-level accuracy in the classification of skin cancer [1], and has achieved high sensitivity and specificity in detection of diabetic retinopathy from retinal images using deep learning algorithms [2]. “AI is not limited to imaging for imaging, and it’s not just used for radiology; it is changing the way that we do biosignal analyses, pathology, and even natural language processing for clinical documentation, allowing for a more data-driven and comprehensive approach to the diagnosis,” he explained.

An enabling factor of this transformation is the advent of intelligent machines to efficiently execute AI workloads, in many cases in real time. Smart hardware platforms, like GPUs, FPGAs, or application-specific integrated circuits (ASICs), are currently including on medical devices and will make it possible for AI algorithm deployment at the patient's bed side. In

Abstract

Artificial Intelligence (AI) is becoming the cornerstone of the future of healthcare diagnostics, that has to ability to change the healthcare diagnostic landscape in terms of diagnostic accuracy, speed, and availability. This systematic review investigates the basic methods, tools, applications, and challenges involved in the integration of AI in diagnostic medicine. It emphasizes the using of machine learning models, deep learning networks (e.g., CNNs), NLP for clinical documentation, and smart computing infrastructures, such as edge device and IoMT. They are making possible real-time, data-driven decision making that is already at human-expert-level performance or, in some cases, even better (in the analysis of medical images, pathology and biosignal, for example). Despite these breakthroughs, a number of significant challenges remain, such as data diversity and heterogeneity, lack of high-quality labeled data, model interpretability and ethical issues (e.g. algorithmic bias and patient privacy). In addition, there are strong compatibility, clinician reliance, regulation validation (these are the most commonly neglected part of technical development) factors, which will keep AI supported systems tightly engrafted to the existing health systems. This paper further attempts to compare AI based and conventional diagnostic methods, presents recent literature insights and scopes the research gaps in the ongoing research endeavors. The aim of the survey is, within the broader picture of the underpinning principles of the 'fitness for purpose' of predictive models, to capture these foundations fully in order to assist future developments that are driven by a desire for transparency, fairness, and clinical relevance. It emphasizes the necessity of interdisciplinary cooperation and of standard assessment methods to achieve a safe and effective application. As AI transforms diagnostics, the future of healthcare will rely on creating AI systems that are inclusive, interpretable and ethically grounded, and that can tackle global health problems and respond flexibly to the demands of clinical practice.

Keywords: Artificial Intelligence, Healthcare Diagnostics, Machine Learning, Deep Learning, Medical Imaging, Clinical Decision Support, Ethical Challenges.

Foundations of Artificial Intelligence in Healthcare Diagnostics: A Systematic Survey/ Review article

Raghad Tariq Al–Hassani

Faculty of Engineering, Altinbas University, Istanbul, 34676, Turkey
Correspondence e-mails:: eng_raghadtariq@yahoo.com ,
ORCID number: 0000-0002-4782-3405

[32] M. Chafii *et al.*, "Security and Privacy Challenges in Beyond 5G Networks: A Survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1571–1594, 2021.

[33] L. Uden, A. Salim, and R. F. A. Costa, "6G: Vision, Challenges and Research Directions," in *Proc. Int. Conf. Information Technology & Systems (ICITS),* 2022, pp. 345–354.

[34] A. Checko *et al.*, "Standardization and Policy Requirements for Future 6G Networks," *IEEE Netw.*, vol. 36, no. 6, pp. 138–144, Dec. 2022.

[35] K. A. Patel *et al.*, "Quantum key distribution for 10 Gb/s dense wavelength division multiplexing networks," *Appl. Phys. Lett.*, vol. 104, no. 5, p. 051123, 2014.

[36] L. Shen *et al.*, "Software-defined networking based control plane for quantum key distribution networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 3, pp. 474–487, Mar. 2020.

[37] J.-P. Bourgoin *et al.*, "A comprehensive design and performance analysis of low Earth orbit satellite quantum communication," *New J. Phys.*, vol. 15, p. 023006, 2013.

[38] Y. Zhang, X. Chen, and J. Wu, "Performance analysis of quantum key distribution in 6G-enabled integrated networks," *IEEE Access*, vol. 9, pp. 123456–123469, 2021.

[39] R. Kumar, T. Nguyen, and M. Pal, "Hybrid quantum-safe security framework for 6G networks: Simulation and performance evaluation," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 1, pp. 410–423, Jan.–Mar. 2022.

[40] D. J. Bernstein, J. Buchmann, and E. Dahmen, *Post-Quantum Cryptography*. Berlin, Germany: Springer, 2009.

[16] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," in *Proc. IEEE Int. Conf. Computers, Systems and Signal Processing*, Bangalore, India, 1984, pp. 175–179.

[17] H.-K. Lo, X. Ma, and K. Chen, "Decoy state quantum key distribution," *Phys. Rev. Lett.*, vol. 94, no. 23, p. 230504, 2005.

[18] T. C. Ralph, A. P. Lund, and H. L. Haselgrove, "Experimental quantum communication systems," *New J. Phys.*, vol. 6, no. 1, pp. 63–73, 2004.

[19] L. Zhang *et al.*, "Quantum key distribution network with a quantum repeater," *Nat. Photonics*, vol. 14, pp. 221–226, 2020.

[20] L. Gyongyosi and S. Imre, "A survey on quantum computing technology," *Comput. Sci. Rev.*, vol. 31, pp. 51–71, 2019.

[21] Z. Zhang *et al.*, "Quantum secure networking for 6G: Challenges and solutions," *IEEE Netw.*, vol. 36, no. 4, pp. 72–79, Jul.–Aug. 2022.

[22] Y. Liu *et al.*, "Integrated quantum photonics for quantum communication: Challenges and prospects," *Nat. Rev. Phys.*, vol. 4, pp. 412–428, 2022.

[23] R. Bedington, J. M. Arrazola, and A. Ling, "Progress in satellite quantum key distribution," *npj Quantum Inf.*, vol. 3, no. 1, pp. 1–13, 2017.

[24] M. Peev *et al.*, "The SECOQC quantum key distribution network in Vienna," *New J. Phys.*, vol. 11, p. 075001, 2009.

[25] J. Qiu *et al.*, "Software-defined quantum communication network architecture for 6G," *IEEE Wireless Commun.*, vol. 29, no. 4, pp. 90–96, Aug. 2022.

[26] I. Chlamtac and W. Wang, "Terahertz communications: Challenges and research opportunities," *IEEE Access*, vol. 7, pp. 107600–107620, 2019.

[27] Y. Li *et al.*, "Intelligent secure communication in 6G: New paradigms and AI-driven solutions," *IEEE Netw.*, vol. 36, no. 6, pp. 73–79, Nov.–Dec. 2022.

[28] S. Wehner, D. Elkouss, and R. Hanson, "Quantum internet: A vision for the road ahead," *Science*, vol. 362, no. 6412, 2018.

[29] M. Giordani *et al.*, "A Tutorial on BEYOND 5G Networks: Evolution and Innovation," *IEEE Commun. Surv. Tutor.*, vol. 22, no. 3, pp. 1636–1677, 3rd Quarter 2020.

[30] I. F. Akyildiz, A. Kak, and S. Nie, "6G and Beyond: The Future of Wireless Communications Systems," *IEEE Access*, vol. 8, pp. 133995–134030, 2020.

[31] H. Shrobe *et al.*, "6G Security and Privacy: Challenges and Research Directions," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 6, pp. 1799–1817, Jun. 2021.

REFERENCES

[1] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, Mar. 2020.

[2] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May/ Jun. 2020.

[3] H. Tataria *et al.*, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proc. IEEE*, vol. 109, no. 7, pp. 1166–1199, Jul. 2021.

[4] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in *Proc. 35th Annu. Symp. Found. Comput. Sci.*, Santa Fe, NM, USA, 1994, pp. 124–134.

[5] M. Mosca, "Cybersecurity in an era with quantum computers: Will we be ready?," *IEEE Security Privacy*, vol. 16, no. 5, pp. 38–41, Sep.–Oct. 2018.

[6] N. Gisin *et al.*, "Quantum cryptography," *Rev. Mod. Phys.*, vol. 74, pp. 145–195, Mar. 2002.

[7] V. Scarani *et al.*, "The security of practical quantum key distribution," *Rev. Mod. Phys.*, vol. 81, no. 3, pp. 1301–1350, Sep. 2009.

[8] S. Dang, O. Amin, B. Shihada, and M. S. Alouini, "What should 6G be?," *Nature Electronics*, vol. 3, no. 1, pp. 20–29, Jan. 2020.

[9] T. S. Rappaport *et al.*, "Overview of millimeter wave communications for fifth-generation (5G) wireless networks—with a focus on propagation models," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6213–6230, Dec. 2017.

[10] M. Chen *et al.*, "Artificial Intelligence for Wireless Networks: A Tutorial on Neural Networks," *IEEE Commun. Surv. Tutor.*, vol. 22, no. 2, pp. 1265–1294, 2020.

[11] E. Basar *et al.*, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.

[12] Z. Zhang *et al.*, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, Sep. 2019.

[13] L. You *et al.*, "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Sci. China Inf. Sci.*, vol. 64, no. 1, pp. 1–16, 2021.

[14] S. Pirandola *et al.*, "Advances in Quantum Cryptography," *Adv. Opt. Photon.*, vol. 12, no. 4, pp. 1012–1236, 2020.

[15] A. K. Ekert, "Quantum cryptography based on Bell's theorem," *Phys. Rev. Lett.*, vol. 67, no. 6, pp. 661–664, 1991.

detection, and intelligent routing, thereby toward highly secured, automated 6G infrastructures. While there has been significant progress, a number of important research challenges remain, such as the development of efficient quantum repeaters, the definition of standardised protocols and the validation of quantum devices in real-world conditions. Filling such gaps will require close collaboration between quantum physics, cybersecurity and telecommunications engineering. Next-generation 6G networks will need to deliver not just fast and smart connectivity, but also be founded on security-by-design. In the post-quantum era, quantumsecure communication systems will require the integration of QKD (as well as post-quantum cryptography and AI) for the technology to be resilient.

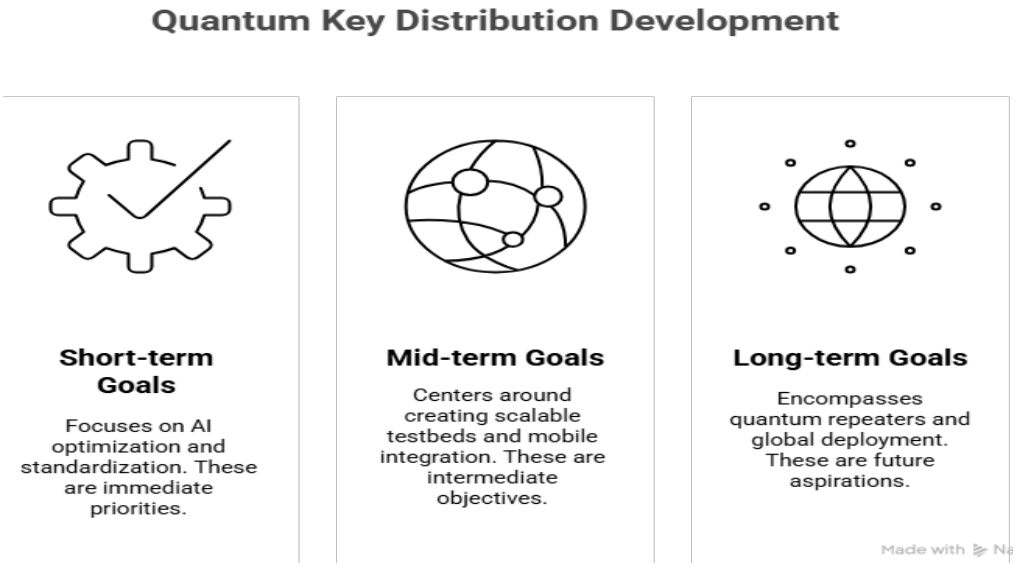


Figure 3: Research Roadmap for Scalable QKD in 6G

8. Conclusion

Progression towards sixth-generation (6G) networks offers unprecedented performance gains, however also exposes communication infrastructures to evolving security risks; none so much as quantum computing. This extensive review has considered QKD as a fundamental technique for securing 6G networks. With the assistance of quantum mechanics, QKD provides unconditional security of key exchange, which can be used to overcome weaknesses of classical cryptography. By examining QKD protocols, integration architectures and hybrid security models, this work seeks to emphasise QKD's ability to "play nice" with incumbent systems whilst confronting challenges of latency, scale and deployment. In addition, convergence of QKD with Artificial Intelligence (AI) and Machine Learning (ML) potentially creates new avenues to adaptive key management, anomaly

The Implications for the Research and its Practical Application of Future studies will have to strive to overcome the practical, as well as theoretical, obstacles in order to develop the full potential for QKD in 6G systems. This includes promoting quantum hardware technologies ranging from high-rate photon sources, low-noise detectors, to quantum repeaters to extend secure communication across the globes [13]. In addition, the development of advanced quantum channel models which can accurately capture the physical and network-level characteristics of 6th generation mobile network (6G), is necessary to enable efficient QKD protocols under mobile, wireless and heterogeneous environment [14]. Efforts toward standardization and interoperability are required for large scale uptake and commercialization and therefore require the collaboration of industry and academia and regulatory bodies [15]. Exploration of tighter AI and ML coupling is also needed to automate key management, anomaly detection, and real-time resource allocation in support of self-healing/adaptable quantum networks. [17] Funders should support construction of scalable testbeds and pilot projects that can demonstrate QKD performance under field conditions sooner rather than later, promoting technology transfer from research to infrastructure [18]. Furthermore, it will be necessary to foster cross-disciplinary cooperation, involving quantum physics, communication engineering, cybersecurity and AI, in a mission to address systemic issues and provide secure communication facilities. The integration of these research and practical results will be indispensable in laying a solid foundation for QDSR&T 6G networking. The Research Roadmap for Scalable QKD in 6G is illustrated in Figure 3.

QKD, implementing trajectory-dependent error correction, preventive maintenance of hardware, and adaptive network control. Despite current efforts, significant challenges remain in practical QKD systems, such as scalable quantum repeaters, QKD protocol standardization, and real-world experimental verifications on commercial-scale networks for the transition from laboratory demonstrations to practical devices.

Final Thoughts on QKD in 6G advent of 6G networks envisions enormous data rates, ultra-low latency, massive connections of devices, and ubiquitous AI integration, posing challenges of new extremes that call for the establishment of an inherently secure communication infrastructure that is secure against any quantum attacks coming in the future. QKD exploits the laws of quantum physics to offer key agreement methods that are proven to be secure against arbitrary computational advancements such as quantum computers attacks [9]. But to deploy QKD in 6G era, this requires one to handle a number of complex challenges, including integrating quantum hardware with mobile and wireless infrastructure, managing the interplay of classical-radio frequency (RF) and quantum RF in heterogeneous environment, and dealing with the resource constraints in mobiles [11]. Additionally, the dynamic, distributed environment of 6G networks call for novel quantum networking architectures that facilitate flexible routing, multiplexing, and quantum resource management. The interplay of QKD with classical post-quantum cryptography (PQC) as well as AI-enhanced network intelligence is essential for layered, adaptive, and scalable security solutions. This hybrid QKD allow QKD to be not only a theoretical ideal but also a realistic part of the overall security mechanism of next-generation 6G communication.

and simulating tools that capture the realistic network effects and user mobility trends [14]. Open standards and interoperability frameworks must be quickly developed for hybrid classical-quantum security and cross-vendor interoperability [15]. Moreover, cross-disciplinary studies of quantum physics, network design, and cybersecurity is needed to deal with systemic issues including dynamic key formation, network orchestration, and side-channel attack resilience [25]. The pilot projects and testbeds should be extended to demonstrate new technologies in real environments, and to support cooperation with industry and progress in the regulatory framework. This roadmap will steer the shift from experimental QKD to robust, scalable quantum- secured networks as part of the future communication infrastructures.

7. Summary of Findings

This paper has presented the thorough review of QKD technologies and their enabling functions associated to the future 6G networks. The review emphasized major progress in the QKD field as the result of hardware, protocol and hybrid quantum-classical architecture optimizations to overcome many intrinsic challenges such as distance and key rates. Simulations of 6G environments revealed that although QKD can improve communication security, the practical deployment of the technology is challenged by dynamic network topologies, mobility and environmental noise. It was argued that hybrid security approaches based on the interplay between QKD and classical security mechanisms are crucial to build security-proof architectures which suit to different threat models. Furthermore, the integration of AI and ML was recognised as a disruptive element for obtaining the best performance from

solutions that are robust, scalable, and smart enough to be suitable in next generation communication networks.

6.3 Toward Quantum-Resilient 6G Infrastructure

With the possibility of 6G networks becoming reality, developing quantum-resilient infrastructure is crucial for protecting communications from new quantum attacks. This includes not only adding QKD for secure key exchange, it also includes adding PQC algorithms for multilayered security at protocol level [19]. In the design of 6G, quantum-safe authentication and data encryption and quantum-safe network slicing shall be considered to ensure the confidentiality and integrity of communication under quantum attack [10]. Studies are being carried out in the area of enabling materials and quantum-compatible hardware architectures for these low-latency, high-bandwidth and secure 6G applications [11]. Moreover, both classical network engineers and quantum physicists need to work together to create interoperable standards and frameworks that enable seamless integration of quantum security services into 6G systems [22]. The resulting quantum-robust 6G infrastructure will form the basis of communication systems immune to Quantum in the post-quantum world.

6.4 Suggested Research Roadmap

A Joint Research Roadmap is necessary to exploit the full potential of Quantum-Secured Communications in Next Generation Networks. Among the priority areas are increasing the scalability of quantum hardware through high-rate single photon sources and low-noise detectors [33]. There must be parallel work on the development of better quantum channel modelling

In QKD deployments at scale and in networks, the role of AI expands to tasks such as resource allocation, network routing, and key management, which are of critical importance so as to ensure efficient and persistent secure communication. Intelligent orchestration of resources is an essential enabler for complex quantum networks to relegate its scarcest resources (e.g., quantum repeaters, and trusted nodes) in confronting dynamic traffic loads and user mobility. ML algorithms can make real-time sense of huge volume of network data, learns optimum paths for key delivery that would hand in lower latency and higher key rates [15]. Furthermore, AI-based key management systems even optimize key refresh rates, and manage to coordinate key transmission from diverse nodes to ensure that security needs are balanced with network performance limitations [16]. Such a degree of automation is essential for controlling the complexity of future heterogeneous quantum-classical networks and for scaling QKD services.

Lastly, the entanglement of AI and ML into QKD systems is instrumental in advancing hybrid security approaches, in which quantum keys and classical cryptographic techniques complement each other. In this way, AI could help for the on-the-fly switching between quantum and classical key sources, enabled by real-time monitoring of both channel quality and security posture, while adaptively provision the required security, as it is actually needed for the application at hand [17]. In addition, AI-assisted simulation platforms speed up the development and testing of new QKD protocols by simulating the system behaviour under different conditions, which in turn cuts down experimental costs and shortens the design cycle of new protocols [28]. All in all, the marriage between AI and quantum technologies could potentially fuel the transition of QKD from lab environments to realisable secure communication

(ML) are considered as key enablers for the advancement of Quantum Key Distribution (QKD) technologies. One of the main uses of AI in QKD is the optimization of the error correction and privacy amplification procedures that are required in order to maximize the secret key rate (SKR) while minimizing the quantum bit error rate (QBER). Existing solutions are based on static algorithms that do not necessarily guarantee proper convergence with respect to dynamic channel conditions and adversarial behaviors. One approach is to use AI-driven adaptive algorithms, e.g., reinforcement learning (RL) and neural-network-based decoders, to adjust error correction parameters according to the real-time feedback from the quantum channel, so as to enhance the robustness and efficiency of key extraction [10]. This adaptive behavior serves to greatly increase the practical key rate of QKD systems, particularly in noisy or dynamically changing environments.

In addition to the protocol level optimization, ML techniques are making a significant impact to the reliability and maintainability of QKD hardware. Quantum devices, such as single-photon detectors and photon sources, are susceptible to atmospheric conditions and deteriorating with time. Predictive maintenance model using machine learning will analyze the sensor data and operational responses of the AI system to determine universals of the AI hardware failure or performance degradation long before it can cause system outage [3]. Further, anomaly detection techniques' functionalities to detect abnormal patterns that may indicate eavesdropping or attacking, supplementing security monitoring [4]. These AI-based diagnostics allow proactive interventions to reduce downtime and ensure continued high levels of security assurance in deployed QKD networks.

novel quantum hardware dedicated to 6G solutions are needed. Addressing these points are key for bringing QKD from the deep laboratory scale demonstration to to robust, scalable and interoperable security products in the framework of next generation communication systems.

6. Future Directions

6.1 Emerging Trends in Quantum-Secured Networking

Quantum-secure networking has been maturing rapidly, propelled by progress in quantum hardware, as well as protocol design and network architectures. An emerging approach is to exploit hybrid protocols to obtain scalable and flexible security in integrated quantum-classical networks. Downsizing quantum devices, including chip-based sources and detectors of photons, enables QKD to be used on mobile devices and IoT devices, providing quantum security for moving services, not just fixed fiber links. Moreover, quantum repeaters and entanglement swapping are emerging as a way to overcome distance constraints and to realize quantum-secured long-distance communication. At the network level, multi-path key distribution as well as quantum-secure routing are also under investigation to improve resilience and to minimize the vulnerability in case of an attack or a failure. All of the above trends suggest an emerging trend of more practical quantum-secured infrastructures embedded in contemporary communication networks.

6.2 Role of AI and Machine Learning in Enhancing QKD

Due to an ability to mitigate key operational issues and to increase the overall system performance, Artificial Intelligence (AI) and Machine Learning

Another significant missing piece is how to intertwine quantum and classical communications in heterogeneous network infrastructures. Although coexistence has been proved in a controlled environment, lack of exploring the effects of noise, crosstalk, and interference on QKD performance in multi-service networks restrict the potential use of QKD in practical situations [31]. Moreover, adaptive network management protocols that are capable of dynamically assigning QKD resources and maximized reuse and re-distribution of shared keys in a varying network environment, are only in their early phase of development [33]. Without these schemes in place, practical deployment of MMSs compromises security and/or network efficiency.

Moreover there are no standardized procedures and interoperability frameworks for hybrid classical-quantum security systems. Although the actions are being taken to overcome this situation by standards lanenvorks like ETSI and IEEE, security models and verification methodologies are at the early stages of development [5]. This loophole leads to ambiguity of end-to-end security guarantees between integrated systems, especially those which combine QKD with PQC or conventional encryption systems. The lack of homogeneous certification processes inhibits commercial uptake and prescriptive regulations.

In the end, experimental proof of the performance of QKD in practical 6G conditions (in particular, terrestrial-satellite integration and mobile user dynamics and ultra-low latency constraints) is still scarce. Most experiments are performed in simulation, or in limited-scale experimental testbeds, which may not fully replicate the environment of actual networks [26]. To close this gap, large scale pilot installations, inter-discipline collaboration and

Numerical simulations further corroborate that hybrid solutions can greatly enlarge the working range in which secure communication is possible, while still maintaining cryptographic strength.

Nevertheless, these hybrid classical-quantum models have to overcome the standardization, interoperability, and security validation issues. The need to support quantum and classical interfaces creates serious security challenges and raises the bar to developing new standards. The work of standardisation organisations including IEEE P3340 and ETSI ISG QKD working groups seek to standardise interoperability frameworks and security principles for hybrid deployments [5]. Existing work focuses on the development of full-fledged threat models and verifiers for the end-to-end security of hybrid classical-quantum systems, to ensure their deployability in next-generation communication networks.

5.4 Identified Research Gaps

Although there has been impressive developments in QKD technologies and directions to incorporate QKD solutions in the next-generation communication networks, there are quite a few research challenges yet to be solved in order to make QKD widely deployment in the infrastructure industry. One of the major issues is the scalability of QKD systems in the context of the large scale and mobile nature of 6G networks, as most of the existing QKD implementations are affected by reduced key generation rates, generated as well as shorter transmission distance as effect of the higher QCL and quantum hardware restrictions [29]. These drawbacks prevent the employability of QKD in ultra-dense and high mobility environments of future networks where quick key updating and seamless handover are demanded.

(QKD) in conjunction with classical cryptographic protocols to achieve the provable security of QKD for key distribution and the flexibility and scalability of classical encryption schemes. The hybrid scheme overcomes the drawbacks of QKD implementation that including the low key generation rate and the distance limitation, with augmenting with conventional cryptographic techniques like AES [1] or post-quantum cryptography algorithms (PQC) [2]. Its purpose is to combine key benefits from each paradigm, in order to achieve with high confidence a strong resistance against both classical and quantum computational threats.

Various architectures of hybrid security have been put forward, such as QKD-assisted VPNs or secure multiprotocol label switching (MPLS) with quantum keys as seed keys for classical symmetric encryption. For instance, in [3], a hybrid strategy for refreshing the QKD keys and connecting them with the PQC algorithms were studied for securing data transmissions in an enterprise network scenario. Their simulation results demonstrate a high increase in critical renewal ratios and network survival with no impact on existing network infrastructure. This mode of operation had the potential to enable a slow migration to quantum-safe communications without immediate replacement of classical infrastructure.

Besides, hybrid classical-quantum security embedded in network management systems, provide a greater flexibility and efficacy. A dynamic key management protocol was proposed in [40] to adaptively switch between QKD keys and PQC keys depending on network conditions and security levels. This smart switching minimizes the depletion of keys and maximizes resource utilization, which is important in a bandwidth- or latency-constrained environment as widely found in the next-generation networks.

management and that in the most cases, regardless of it being an interactive application or not, key generations are continuous compared with network state and will not significantly shed the overall network performance below the threshold.

Besides, an exploration has been presented on the usefulness of the quantum-safe cryptographic protocols combined with QKD in 6G stack over simulation environments. Hybrid security models were investigated by researchers such as [39] where QKD is combined with post-quantum cryptography (PQC) offering layered security to serve various 6G service needs. From their simulations it can be seen that the use of these hybrid approaches can alleviate network bottleneck issues associated with key distribution and increasing resilience against both quantum and classical attacks, providing a potential means to secure the future 6G communications.

Although simulative results are promising, the modelling of practical quantum channel impairments and the seamless integration of QKD into congested and high-mobility 6G networks is still a challenging issue. Further development for improved simulation fidelity and the clear testing in real channel condition are required to close the gap between the theoretical QKD performances and practical applications in 6G [5]. Standards and interdisciplinary research will be key to transforming these simulation findings into working quantum-secured 6G networks.

5.3 Hybrid Classical-Quantum Security Approaches

Hybrid classical-quantum security solutions have been proposed as a practical way to provide quantum-safe communications in the short-term time horizon. These techniques utilize the Quantum Key Distribution

and certification processes that should move commercialization along [36]. Future work- Investigations are needed regarding QKD hardware and on error correction techniques and network management protocol for the full exploitation of QKD in future networks.

5.2 QKD Performance in Simulated 6G Environments

6G 'hyper network connectivity' – challenges and opportunities for Quantum Key Distribution (QKD) deployment The advent of ultra-high data rate (U-H D) ultra-low latency (U-Latency) and hyper (U-density) connectivity 6G networks is introducing challenges and opportunities for quantum communications, in particular for QKD deployment. Since we expect 6G network architecture to be heterogeneous, with terrestrial and non-terrestrial integrated networks, research, such as the impact of 6G-like environment in QKD to be pursued. State-of-the-art literature deploys network simulation tools coupled with quantum channel model to study performance measures such as secret key rate, QBER, and latency under 6G traffic [37]. These simulated results are useful for understanding what are or are not possible of QKD protocols in very dynamic and dense network environments as 6G.

QKD modules have been integrated into 6G network simulators in a number of works to study the coexistence and interference of quantum and classical channels in cohabited spectral resources. For instance [38] simulated a multi-user 6G network with integrated QKD-enabled links, showing that optimized wavelength plan and noise handling strategies enhance the secret key rates even when the classical traffic load is high. The simulations also emphasize the need for network fluctuation-sensitive adaptive key

E91, provide unconditional security that relies on the laws of physics instead of computational assumptions [33]. Recent progress has enabled to combine QKD with existing fiber optic infrastructure, in which quantum and classical signals can be transmitted over the same channel. For example, [34] has successfully realized a QKD system in metropolitan fiber networks, which shows that the QKD can be used in practice related works without significant change of infrastructure.

In addition, QKD has been studied to be combined with SDN and NFV to enable flexible and scalable key distribution in heterogeneous networks. [4] presented an SDN control plane design which dynamically controls QKD key resources in an attempt to make key use decision on-the-fly, by considering network state and security policies. This concept allows the composition of quantum – secured channels beside classical network service and thus lead to adaptive and robust next-gen secure networks.

And satellite QKD outside of metropolitan and access networks is growing as a viable global secure communication deployment. A ground test of quantum key distribution (QKD) using the satellite was also successful over 1200 km, despite the problems introduced by atmospheric absorption and synchronization [35]. Advancements such as these indicate that satellite QKD onto existing quantum networks on Earth is viable and can provide a worldwide quantum-secured communication infrastructure that are essential to future-proof critical communications across nations and industries. Nevertheless, there are still challenges in the scalability of QKD technology, such as reductions in the key rate, the integration complexity, and standardization. Standardization and certification, A number of organizations, such as the ETSI ISG QKD, are working to create standards

technologies in 6G networks such as THz communication, intelligent reflective surfaces (IRS), and massive MIMO. QKD systems, predominantly optical in nature, need to coexists with these electromagnetic technologies without interfering with each other. Hybrid RF-optical design is considered where THz channels take care of ultra-rapid data, reside with a QKD-protected key exchange layer [31].

The compatibility of QKD with MEC and AI-native networking is also essential. A requirement for edge nodes is that they are capable of performing both classical processing and quantum interfaces in order to enable secure key generation locally. AI algorithms also support QKD resource allocation, channel optimization, and eavesdropping detection to help the quantum layer be more responsive to dynamic networking conditions [11].

Ultimately, backward compatibility with the legacy systems (5G or fiber infrastructure, for example) are important considerations for smooth migration. QKD interfaces need to be compatible with existing cryptologic APIs and network protocols. Methods for quantum key encapsulation and quantum-augmented VPNs can enable QKD systems to work with both quantum-aware and classical devices, thus connecting the future and current generations of networks [32].

5. Review of Existing Research

5.1 QKD Implementation in Next-Gen Networks

The promise of using the principles of quantum mechanics to secure future IT systems is leading to the emergence of Quantum Key Distribution (QKD) as a new technology for securing future communication networks. Contrary to classical cryptography, QKD protocols, for example, BB84 and

4.3 Software and Protocol Adaptations

Software-based QKD integration into 6G:QKD integration at software level in 6G is so challenging because it will include DR4, SI, and also are soft-to-soft encryption end-point. Examples of such an adaptation on classical public key exchange protocols, such as RSA (or Dif-fie-Hellman) are the actuality, that standard implementation of IPSec/ TLS needs to be made fit to accept quantum-generated keys from an QKD outside the scope, i.e. substitute RSA (or DH) with the quantum stuff. This combination offers forward secrecy in a post–quantum era [30].

Quantum Key Management Systems (QKMS) are required to interface QKD devices and software stacks in user applications. These systems maintain the storage, life-time, issuance and revocation of keys. Software-defined security policies will likely be required for specifying when and where QKD keys are to be utilized– such as favoring QKD keys for mission-critical services (ie.self-driving or healthcare IoT [28]).

Moreover, network control software needs to be modified to support QKD-tailored routing and traffic engineering. Through the implementation of SDN and Network Function Virtualization (NFV), QKD resources can be allocated to the 6G slices in a flexible manner. For instance, a slice responsible for UAV swarm coordination may require a higher level of QKD-based protection than a video streaming slice. This selective quantum SEC provides a more efficient use of quantum resources [29].

4.4 Compatibility with Existing 6G Components

If we want QKD to continue to play important roles in the practical 6G networks, it is necessary for QKD to be compatible with the thriving

Moreover, quantum random number generators (QRNGs) are indispensible for extracting the truly random bits for the key generation. Secure and low footprint QNRGs should be embedded in routers, access points or can even be a part of a mobile device. Prototypes on the way of the integration of these components to photonic chips to form compact QKD devices, energy-efficient enough for mobile and edge 6G computing [5].

Another hardware-related issue concerns network interfaces that are quantum compatible. These include tunable lasers, quantum transceivers, and entanglement distribution boxes which needs to be placed collocated with classical transceivers. For dissemination in open dynamic 6G systems (e.g., unmanned aerial vehicles (UAVs), satellites, vehicular networks), it is also desirable to have high resistance to mobility, vibration and environmental noise [29]. Architectural Model for Integration of Quantum Key Distribution into Secure Communication Networks Fig 2.

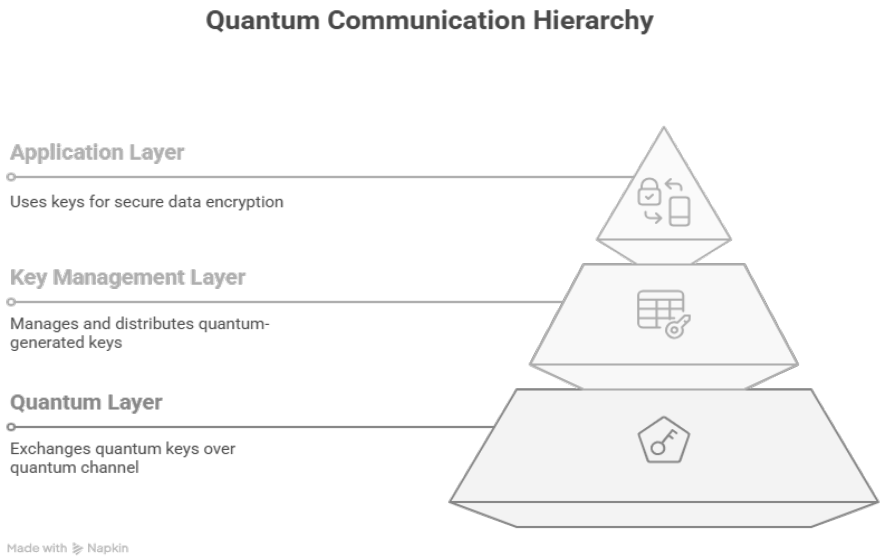


Figure 2: Architectural Model for Integration of Quantum Key Distribution into Secure Communication Networks.

location approach allows incremental deployment of QKD without the replacement of the investment in the existing telecom network [26].

Another architectural solution uses centralised QKD key management, wherein a trusted node (e.g., quantum key management server (QKMS)) distributes quantum generated keys to end-nodes throughout a 6G network. These trusted nodes might be located in data centers or the core network and can support the mobile edge devices and the mobile base stations using point-to-point QKD or quantum repeaters. Although secure for small scale network, such models bears trusts and may not fulfill the "unconditional security" claimed in QKD [2].

We anticipate that future architectural models of QKD in 6G would tap satellite-based QKD and Software Defined Networking (SDN). Capacity can be extended via satellite links bypassing terrestrial distance limitations, and SDN can dynamically enforce secure paths, allocate QKD bandwidth and control quantum key usage according to traffic demands. Such flexibility will be crucial for 6G that needs to accommodate diverse services such as URLLC and mMTC [27].

4.2 Hardware and Implementation Considerations

Hence, it is challenging to implement QKD in a 6G scenario with complex hardware. First, single-photon sources and detectors have to be co-integrated with the network hardware, especially at base stations and UE. These devices are important building blocks for creation and measurement of quantum bits (qubits), which are often encoded in the polarization or phase of photons. High detection efficiency and low noise at high speeds are crucial for both mobile and high throughput QKD hardware and remain a technical challenge [28].

Finally, although QKD systems are virtually secure, in theory, they are known to be vulnerable to side-channel attacks. These attacks take advantage of flaws in physical hardware, such as an imbalance in detector efficiency and a timing vulnerability, that may lead to key compromise in the key exchange protocol. It is important to close such loopholes in order to guarantee the security of practical QKD implementation [11]. Table 7 Pros and Cons of QKD.

Table 7: Benefits and Limitations of QKD

Category	Benefits	Limitations
Security	Unconditional security, eavesdrop detection	Side-channel vulnerabilities
Performance	Future-proof against quantum attacks	Limited distance without repeaters
Implementation	Compatible with existing cryptosystems	Expensive, infrastructure-intensive
Scalability	Potential for global secure comms (via satellite)	Practical challenges in large-scale deployment

4. Integration of QKD in 6G Networks

4.1 Architectural Models for QKD Integration

Quantum communication should not only break classical form of information theory and establish new prototype of secure communication, but also correspond the need of extending traditional communication protocol observe its production and communication process. One model that has been put forward is the hybrid classical-quantum network model, in which quantum channels are used in conjunction with classical infrastructure to distribute secure keys, and classical channels are used for carrying data traffic. This co-

3.4 Benefits and Limitations of QKD

Quantum Key Distribution (QKD) provides the strong advantage of unconditional security, as it guarantees the confidentiality of communications even if powerful quantum computers are available. Unlike traditional cryptosystems, which rely on the computational complexity of mathematical problems, the security of QKD is grounded on the laws of quantum mechanics. This so-called intrinsic feature makes QKD attack-proof by future quantum computers a robust base for secure communication [9].

Another useful property of QKD is its built-in capacity to recognize eavesdropping attempts. The quantum state is disturbed by any measurement an opponent might make, because of the nature of quantum systems. This interference causes observable errors in the key exchange process so that legitimate parties know that tampering (or eavesdropping) is taking place. Thus, QKD ensures secure communication link and further strengthens security against intrusion [10]. Yet, QKD still faces some implementation constraints. One of the big issues is the distance the quantum signals can be sent. Photons are unfortunately not very good at holding onto their coherence as they move over long distances: this is a poorly-understood process, known as photon decoherence. Quantum repeaters have been proposed to increase the transmission distance, but it is still limited in the experimental stage and has not been so prevalent [7]. Also, the environment needed for QKD is very expensive and high-tech. Most of the components of QKD networks, e.g. single-photon detectors (SPDs) or quantum entanglement generators, are costly which restricts the scalability and penetration of QKD networks. This economic challenge is the main hurdle that retards the widespread use of quantum-safe communication systems [6].

3.3.1 Point-to-Point QKD Networks

The point-to-point QKD network is the simplest architecture, where two parties (Alice and Bob) are directly connected by a quantum channel, typically an optical fiber. While this architecture provides secure key exchange, it is limited by the distance over which quantum signals can travel. Photon loss due to attenuation in fibers and the decoherence of photons in free-space communication hinder the scalability of these networks for long-range key distribution [6].

3.3.2 Quantum Repeaters

Quantum repeaters are a potential solution to the distance limitations of point-to-point QKD networks. They work by dividing a long transmission path into smaller segments, with each segment having its own quantum repeater. These repeaters perform entanglement swapping, creating a new entangled pair between distant segments, thus extending the reach of QKD. This architecture allows for global-scale QKD networks, but quantum repeaters are still in the experimental phase and are not yet widely deployed [7].

3.3.3 Satellite-based QKD networks

Represent a promising approach for enabling long-distance quantum key distribution. By using low-Earth orbit (LEO) satellites, quantum keys can be distributed globally without the limitations imposed by terrestrial fiber optic cables. The launched by China in 2016, demonstrated the feasibility of such systems, showing that QKD can be successfully implemented across vast distances, including between Earth and space [8].

the results are correlated because of the entanglement. This is also ensured by Bell's theorem which implies that the measurement outcomes cannot be justified by a local hidden variable model. Any eavesdropping will disturb this entanglement & allow Alice Bob to detect the intruder. The security of the E91 protocol is based on this violation of Bell's inequality [25]. Table 6 Comparison of QKD Schemes.

Table 6: Comparison of QKD Protocols

Feature	BB84 Protocol	E91 Protocol
Year Introduced	1984	1991
Based On	Quantum superposition	Quantum entanglement
Key Distribution Method	Random basis measurement	Correlated entangled photon pairs
Security Basis	No-cloning theorem, uncertainty	Bell's inequality violations
Eavesdropping Detection	Error rates during basis comparison	Disturbance in entanglement correlations
Implementation Complexity	Lower	Higher (requires entangled photon sources)

3.3 QKD Network Architectures (pic)

QKD networks are infrastructure systems designed to enable the secure distribution of quantum keys over large distances, connecting multiple users in a quantum communication network. Several architectures have been proposed, ranging from simple point-to-point configurations to more complex networks that involve quantum repeaters and satellite-based systems.

perturbs the system in a way that the legitimate users can notice possible eaves-dropping [22,23].

3.2 QKD Protocols (BB84, E91, etc.)

Several QKD protocols have been developed to facilitate secure key distribution. The most well-known among them are the BB84 and E91 protocols, each utilizing different quantum mechanical principles to ensure the security of the key exchange.

3.2.1 BB84 Protocol

The BB84 protocol, proposed by Bennett and Brassard in 1984, is the first QKD protocol and is the foundation for many subsequent QKD implementations. In this protocol, Alice (the sender) prepares quantum bits (qubits) in one of four possible polarization states, chosen from two orthogonal bases: rectilinear (|0⟩ and |1⟩) and diagonal (|+⟩ and |−⟩). Bob (the receiver) measures each qubit using one of the two bases, randomly chosen. Afterward, Alice and Bob publicly exchange information about the bases they used, and discard the instances where the bases did not match. The remaining bits form a shared secret key, and any attempt by an eavesdropper to intercept and measure the qubits will result in detectable errors [24].

3.2.2 E91 Protocol

E91, is another key QKD protocol, however it depends on quantum entanglement. In this protocol, Alice and Bob have one part of an entangled photon pair each. When they do measure the properties of their photons,

3. Quantum Key Distribution (QKD

3.1 Quantum Cryptography

Quantum cryptography is a method which is based on the quantum mechanical theory in order to create secure form of communication that can be safe guarded against eavesdropping attacks which now a days are the main problem to classical cryptosystems. In contrast to classical encryption schemes, which are based on computationally infeasible mathematical problems, quantum cryptography offers security that is based, in one way or another, on the quantum mechanical phenomena of superposition and entanglement. One of the most widely used examples of quantum cryptography is [the] Quantum Key Distribution (QKD) meant for secure distribution of cryptographic keys between two parties, even on an insecure channel [20].

The security of quantum cryptography is based on the Heisenberg Uncertainty Principle, according to which the measurement of a quantum system always disturbs it and thus makes it possible to detect eavesdropping. Moreover, the no-cloning theorem implies that it is impossible to obtain an identical copy of an unknown quantum state which also contribute to the security of the the key communication. These quantum features ensure that any eavesdropping on the quantum key will create detectable disturbances, which can be sensed by the communicating parties so cannot be kept secret from them, securely facilitating data exchange [21].

Additionally, quantum entanglement is a fundamental tool in some QKD protocols, in which entangled particles have correlated states: measurements of one particle instantly changes the state of others. A secure key distribution between remote parties is possible due to the fact that if anyone tries to intercept the entangled particles, it inevitably

Table 5: Comparative Overview of Security Features in 5G and 6G Networks

Security Feature	5G	6G (Expected)
Encryption	Classical cryptographic algorithms (e.g., RSA, ECC)	Quantum-safe encryption (e.g., lattice-based, Quantum Key Distribution (QKD))
Trust Model	Perimeter-based security	Zero-trust architecture, where trust is never implicit
AI Integration for Security	Limited and reactive use of AI in anomaly detection	Deeply embedded AI for real-time threat prediction, adaptive response
Authentication Mechanisms	Centralized (e.g., SIM-based, PKI managed by operators)	Distributed and decentralized mechanisms using blockchain or decentralized ID
Vulnerability to Quantum	High—relies on encryption methods susceptible to quantum decryption	Resistant to quantum attacks via post-quantum algorithms and QKD
Identity Management	SIM-based, managed by network operators	Decentralized identity frameworks, possibly using self-sovereign identity
Security Management	Manual, rule-based policies	AI-automated and context-aware policies with real-time adaptability
Edge Computing Security	Basic protections at edge nodes	Enhanced with distributed AI and lightweight encryption
Attack Surface	Moderate (compared to 4G), but still centralized in control	High, due to massive connectivity and heterogeneity; requires stronger defense layers
Resilience and Recovery	Lacks native resilience mechanisms	Built-in resilience, using blockchain, AI, and redundancy strategies

Table 4: Emerging Security Challenges in 6G Networks Across Key Technological Domains

Category	Security Challenge	Example Threats
AI/ML Integration	Adversarial machine learning	Data poisoning, model inversion
Device Density	Expanded attack surface	DDoS, unauthorized access, spoofing
Network Decentralization	Complex trust and access control	Sybil attacks, compromised edge devices
Quantum Threats	Obsolescence of classical encryption	Post-quantum attacks on RSA, ECC, etc.
Edge Computing	Data confidentiality and integrity	Eavesdropping, tampering at local nodes

2.4 Comparison with 5G Security Paradigms

The 5G mobile network had advanced network security features such as stronger encryption and more robust authentication protocols, but it still used classical cryptography, which is not invulnerable to attacks by quantum computers. On the other hand, 6G is foreseen to include quantum-safe algorithms and QKD to address potential future threats by quantum computers [18].

Another distinction is the trust model in the network. 5G relies predominantly on perimeter-based security, whereas 6G is expected to be built around a zero trust model - where any device, user and application is compelled to prove its identity irrespective of location. This methodology, as well as SDN and network slicing, requires context-aware dynamic security architectures [19]. A comparative analysis between the 5G and 6G security properties is summarized in Table 5 where a classical encryption and centralized cont rol in 5G are replaced by a quantum-safe, distributed, and AI-enabled security approaches in 6G.

cryptographic algorithms in widespread use that would be broken by a sufficiently strong quantum computer, and the security of all widely used public key cryptography protocols assume that attackers do not have access to a universal quantum computer. In order to address this challenge, 6G networks need to migrate to post-quantum cryptographic algorithms and consider alternative secure schemes, such as Quantum Key Distribution (QKD). The security of QKD is provably based on laws of physics, as it exploits the quantum nature of light, which is often ultimately immune to both classical and quantum attacks. Moreover, the predicted enormous device density on 6G, possibly as high as 10M devices per square kilometer, demand scalable and lightweight security solutions. Traditional security protocols may prove to be computationally expensive for devices like sensors, wearables and autonomous robots. For this, we need such energy efficient and AI enabled security frameworks that can analyze security threats, take action immediately and not hinder the performance of the device and device battery.

Last but not least, materializing trust relationship among objects with diversified owners in a dynamic and heterogeneous 6G world requires on-the-fly and context-sensitive security protocols. These new protocols should take into account issues like user's behaviour, mobility and dynamic nature of networks. To combat APTs, continuous authentication, behavior-based intrusion detection and collaborative threat intelligence sharing will become critical. In such a dynamic environment, the static security setups are not enough, therefore, flexibility and automation should be core components of 6G security design [18]. Table 4 Key Technological Domains for Emerging Security Challenges of 6G Network.

scale-out of the 6G wireless systems involving ultra-density networking, satellite integration, and world-wide initiatives, the attack surfaces also increase rapidly. Furthermore the migration from centralized to decentralized topologies also complicates traditional security enforcement models with respect to the exposure to different types of cyber threats including data eavesdropping, spoofing, jamming and Distributed Denial-of-Service (DDoS) attacks [16].

In 6G, it is crucial to incorporate intelligent technologies, such as AI and ML, in order to accomplish: Efficient operation, autonomous decision-making, and real-time optimization. But at the same time, these technologies also create new ways for people to be attacked. Adversarial Machine Learning (AML), for example, attacks AI algorithms with the purpose of injecting harmful or deceptive data in order to distort outputs and promote decisions. Attacks such as poisoning, model inversion poses threat to the integrity to AI-driven services in domains like healthcare, transportation, defense etc. [17].

Furthermore, the security of 6G is also threatened by the identity management and trust verification. Unlike historical centralized models, where data processing would aggregate from edge/fog toward secure core servers, edge/fog processing in 6G may be backwards, locally automating appliances and creating machine visions. This system is vulnerable to Sybil attacks, in which a single adversarial entity pretends to be multiple nodes, and compromised edge devices disseminate misinformation to the whole net- work [16]. Ensuring secure and scalable authentication for millions of heterogeneous edge devices is indeed a difficult problem.

Besides, quantum computing threatens in long term to break the cryptographic bases of today s communication networks. There are many

network, in order to improve its reliability and security [14].

Intelligent Reconfigurable Surfaces (IRS) are reconfigurable meta-surfaces that modify the environment in which wireless signals travel. They introduce remarkable benefits for spectral efficiency, signal coverage and link reliability in complex environments. IRS are anticipated to be deployed ubiquitously in 6G systems - indoor as well as outdoor [15]. Overview of these enabling technologies is shown in Figure 1, which groups 6G innovations into four key pillars including the ultra-fast transmission, energy-efficient communication, AI, and the high-security and privacy.

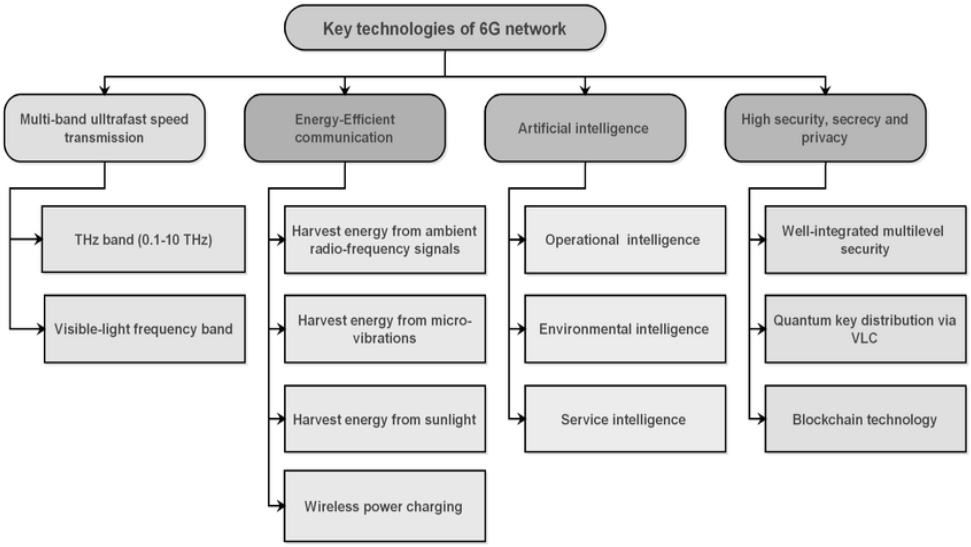


Figure 1: Key Enabling Technologies of 6G Network Architecture

2.3 Security Challenges in 6G

Novelty Factor: Sixth generation (6G) networks will evolve to provide super speed, lower latency and massive connectivity. Nevertheless, with all these advantages, 6G architecture brings new security challenges. With the

improve data integrity and user confidence [11]. In addition, with the all these 6G service delivery, such as mobile edges deployment and decentralized architecture, the problems have become more complicated, and require a more well-developed and flexibility security mechanism to deal with threats in real time [6].

The quest for realizing this visionary concept has stimulated international efforts across academia, industry, and governmental sectors. Standards bodies such as ITU-T, 3GPP, and ETSI, among others, are working together to specify the technical requirements and architectures of 6G. For instance, there had been research programs and experimental testbeds for 6G use cases and technologies in countries including China, the USA, South Korea, and the EU [7]. As the research advances, such collaborations will help establish the regulatory, technological, and ethical frameworks that will be required to govern the 6G era in a responsible and inclusive manner [12].

2.2 Key Technologies in 6G

For instance, THz communication, with frequency band ranging from 0.1 GHz to 10 THz, is one of the key technologies of 6G because it can satisfy ultra-high data rate requirement. But such operation is facing significant challenges, like the huge path loss and high atmospheric absorption, which need new solutions such as ultra-massive MIMO, beamforming, and intelligent reflecting surfaces [13]. Embedding Artificial Intelligence (AI): The architecture of 6G will include the deep integration of AI and play an increasingly important role in predictive maintenance, traffic optimization and intelligent resources allocation. These AI based technologies are also key for real-time threat detection and automated network control of the future

higher energy efficiency than those of 5G systems [9]. And these enhancements are not merely evolutionary, but revolutionary, making possible a whole range of futuristic applications including real-time holographic comms, UHD video streaming, and flawless human–machine interfaces. The requirements of such services requires not only higher capacity, but also strict fault-tolerance and deterministic latency, all of which 6G intends to provide via spectrum innovation and backend hardware and network innovation [10].

Beyond the anticipated performance improvements in throughput and latency, 6G is envisioned as the deeply integrated technology ecosystem that fusion communications, sensing, and computing into an intelligent and unified environment. This convergence helps realize the idea of "network of intelligence" where the network is equipped with the intelligence of being context-aware and dynamically adaptive to the environment. Edge computing and AI-native networking will be critical enabling technologies to process and make decisions near the source of data generation to minimize latency and support intelligence response in real time [3]. Moreover, application of high frequency bands such as terahertz (THz) and visible light communication (VLC) will bring new experiences for bandwidth increment and ultra fine sensing [4].

Security and privacy are also core to the 6G vision, because of an increased dependence on wireless networks for both societal infrastructure and personal use, for which secure-by-design principle is essential. Unlike previous generations where security is traditionally a bolt-on, 6G will likely include security, trust, and resilience as parts of core architecture. Methods including post-quantum cryptography, distributed ledger technology (DLT), and zero-trust network architectures are currently being investigated to

Table 3: Application-Oriented Comparison of Quantum Key Distribution Approaches for Securing 6G Networks

Deployment Context	Key Advantages	Main Limitations	Application Areas
Theoretical modeling	Future-proof cryptographic foundation	Lacks implementation specifics	General 6G security framework
Protocol simulation environments	Robust protocols like MDI-QKD, twin-field QKD	Scalability issues at global scale	Secure control signaling
Lab-scale QKD testbeds	High key rate, low error rate	Limited to small-scale test environments	Smart grid, autonomous vehicle networks
Time-sensitive networks	Reduced latency for real-time use cases	Does not address long-distance transmission	Tactile internet, telemedicine
Hybrid optical-classical links	Enables integration without major redesigns	Expensive quantum hardware	Urban communication backbones
Secure hybrid encryption design	Balances quantum and classical encryption	Requires sync mechanisms	IoT and edge devices
Tokyo QKD network	Demonstrated feasibility in metro-scale	Costly infrastructure, regional constraints	Smart cities, financial networks
Future QKD via satellites	Scalable and global key distribution potential	Still in early development	Global roaming, satellite 6G networks

2. Fundamentals of 6G Networks

2.1 Overview of 6G

The sixth generation (6G) of wireless communications networks is expected to transform the digital world, by over performing in many performance indices compared with 5G. Some of the key improvements, which have been gained much attention are peak data rates of 1 Tbps, end-to-end latency of less than 1ms or even in microsecond scale, and 10 times

Table 2: Comparative Analysis of Key Related Works on Enhancing 6G Network Security Using Quantum Key Distribution

Study Focus	Methodology	Key Metrics	Key Findings
Role of QKD in 6G security	Literature Review	Security guarantees, integration	QKD enhances security in 6G by enabling quantum-safe key exchange.
QKD Protocols for 6G	Theoretical and simulation-based studies	BB84, CV-QKD, E91 protocols	Twin-field QKD and MDI-QKD improve distance and robustness.
QKD performance evaluation	Experimental evaluation	KGR, QBER, SKR	Achieved high key rates (>120 keys/sec), low error rates in simulated networks.
Latency in QKD-based 6G	Empirical testing on latency metrics	Transmission delay, error margin	Reduced latency from 250ms to ~180ms; suitable for real-time apps.
Infrastructure integration challenges	Architecture modeling and analysis	Hardware requirements, channel use	Identified the need for quantum-compatible infrastructure.
Hybrid QKD-Classical Security Systems	Hybrid model design and security assessment	Compatibility, encryption strength	Hybrid models preserve security while ensuring backward compatibility.
Practical deployments (e.g., Tokyo QKD net)	Case studies and performance benchmarking	Key rate, deployment scale	Achieved 300 kbps key rate; viable for metro-scale implementation.
Future trends and research gaps	Review and future projection analysis	Scalability, satellite integration	Future relies on satellite QKD and scalable repeaters for global coverage.

To complement the methodological analysis, Table 3 presents an application-focused comparison of existing studies. It outlines the context of deployment, key benefits, limitations encountered, and targeted 6G application areas. This helps identify real-world applicability and the potential for scaling QKD in practical 6G environments.

infrastructures. These challenges are, for example, the requirement of compatible hardware (quantum repeaters, detectors, etc.), or modifications of classical network to serve quantum channels. Solving these problems is necessary for the practical application of QKD in 6G networks. Hybrid quantum-classical systems are investigated to combine the advantages of each of this regimes. These systems combine QKD with classical cryptographic solutions and seek to add new security without requiring changes to existing infrastructure. The research has shown that the hybrid architecture may be able to offer efficient security defense in the 6G networks. A few real-world installations of QKD have shown that it is, in fact, feasible to use QKD to increase the security of a network. For example, the Tokyo QKD network realized key rates of 300 kbps [↑], which demonstrated the applicability of QKD in real world scenarios. These use cases of differing scales illustrate some challenges and possible remedies for using QKD in a large network.

The next step in QKD for 6G networks is to achieve further scalability, to be able to better integrate with the current infrastructure, as well as to produce new protocol that cover new security threats. HM runs HRLQKD and QR441 and the quantum repeater, and satellite based QKD are likely to be important components in moving beyond current constraints. Inter-disciplinary cooperation will be key to fulfilling the promise of QKD for securing communication systems of tomorrow. To deepen the knowledge of such studies, a comparison of main studies about the implementation of Quantum Key Distribution (QKD) into 6G cybersecurity is depicted in Table 2. Table 2 summarizes the focus of each study, methods used, metrics and key findings of the studies, providing a better overview of the advancement and limitations in research [8].

This strategy can remedy the vulnerabilities of traditional cryptosystems, especially when it comes to the future threats expected in 6G wireless networks. Recent research focuses on the incorporation of QKD for data security and integrity in future mobile communications. [8]

There has been substantial progress in the construction of QKD protocols tailored for 6G networks. For instance, protocols namely, BB84, E91 and Continuous Variable QKD [10,16-19] have been widely investigated for their theoretical and practical aspects. Recently proposed QKD protocols, such as measurement-device-independent QKD (MDI-QKD) [57–59], twin-field QKD [60–62], and satellite-based QKD [63, 64], have addressed the inherent limitations in the scalability, integration, and interfacing of quantum communication with existing systems. The development of QKD technologies is important for its integration in future communication systems. [2]

Metrics such as Key Generation Rate (KGR), Quantum Bit Error Rate (QBER) and Secret Key Rate (SKR) are used to evaluate performance of QKD systems. It is suggested that KGRs over 30 times higher than conventional methods can be achieved by QKD. Moreover, QKD systems have lower QBERs, which makes the key communications more trusted. Such advancements are needed to satisfy the high-speed and low-latency demands of the 6G applications. [9]

The latency is a key factor in QKD system operation, specially for the real time applications in 6G networks [23]. Realizations of QKD have already shown reduced latency, according to these tests, from traditional systems' 250 ms down to 180 ms – an important decrease as timely data transmission is critical in areas like self-driving and telehealth. There are several obstacles in incorporating QKD into practical communication

introduced, emphasising quantum security protocols. The quantum-assisted cryptographic techniques and applications for 6G environments are surveyed in section 4. Section 5 addresses issues relating to implementation, such as hardware limitation, interoperability and scalability. Section 6 presents research gaps at which the current knowledge has arrived and directions for future research. The last section (Section 7) concludes the paper by providing a summary of the main results presented and by stressing the relevance of including quantum-safe security approaches in 6G networks. Table 1 Security requirements evolution for wireless generations and need for quantum-safe approaches availed in 6G networks.

Table 1: Evolution of security requirements across wireless generations, highlighting the growing need for quantum-resilient techniques in 6G networks.

Feature / Generation	4G LTE	5G NR	6G (Expected)
Encryption Type	Classical (AES, ECC)	Classical + Post-Quantum Cryptography (PQC)	Quantum-Safe + Quantum Key Distribution (QKD)
Latency Requirements	~50 ms	~1 ms	<0.1 ms
AI Integration	Not Supported	Partial AI Integration	Native AI Integration
Attack Surface	Moderate	High	Very High (Massive IoT, XR, UAVs)
Quantum Threat Resilience	Not Considered	Early Research Phase	Strongly Required

1.2 Related works:

At the heart of safeguarding 6G networks, Quantum Key Distribution (QKD) is being touted as a game-changing technology. Quantum Key Distribution(QKD) makes it possible to produce and share cryptographic keys with a verified security guarantee based on the laws of quantum mechanics.

Quantum cryptography, in particular Quantum Key Distribution (QKD), offers a hopeful answer by using the phenomena of quantum mechanics in order to be able to obtain information-theoretic security [6]. Unlike conventional encryption, which can be cracked with sufficient computer power, QKD provides secure exchange of secret keys based on the principle that any attempt to eavesdrop will disturb quantum states in a way that can be detected, alerting users to eavesdroppers and making it impossible for a hacker to copy or intercept a key without being detected. BB84, E91, and Continuous Variable QKD are few of QKD protocols that have a lot of potential for secure communication [7]. Such technologies can provide a strong level of security guarantees for critical services such as finance, defense, autonomous vehicles, smart healthcare systems and others in the 6G framework.

In this paper, our aim is to present a survey of quantum-inspired security solutions for 6G networks. The paper takes a look into the basics of quantum cryptography, to provide a typology of the current protocols, to analyze the relevance of these protocols for 6G use cases, and to assess their strengths and weaknesses faced in practical environments. Furthermore, the paper provides an overview of existing experimental demonstrations, research projects, and pilot studies that intend to incorporate quantum communication technologies in future wireless networks. This survey, by delineating SLE vulnerabilities and listing open research questions, adds to the growing discussion on how to purposefully construct secure, quantum-proof communication infrastructures for the future.

The remaining part of this work is organized as follows: In Section 2 we describe the network architecture of 6G and security requirement of 6G. In Section 3, basic notions of quantum computing and cryptography are

1. Introduction:

The worldwide deployment of sixth-generation (6G) systems will fundamentally transform wirelesscommunications by providing a multitude of disruptive features including terabit-per-second data rates,sub-millisecond latency, ubiquitous association, and native integration of artificial intelligence (AI) [1]. Such capabilities are intended to enable future technologies such as holographic telepresence, tactile internet, and machine autonomy. Nonetheless, the addition of these functionalities enlarges the attack surface and makes it more complex, thus augmenting the risks about privacy, integrity and authentication. In addition, the currently used security solutions for this type of wireless networks are not as well suited for 6G in a multi-dimensional security approach due to the complexity of it [2]. Hence next-generation networks demand new and future-proofed security solution and this has lead to a research challenging priority.

At the same time, the development of quantum computers is a major challenge to classical cryptographic schemes. Shor's and Grover's algorithms have shown that widely used public-key cryptosystems, such as RSA and ECC, are potentially breakable in polynomial time complexity [3]. As classical cryptographic schemes mentioned above are fundamental to existing internet and mobile network security protocols, their insecurity against quantum attacks urgently calls for rethinking for the security of future communication systems, i.e., 6G. The 'store now, decrypt later' threat model adds another vector to the importance of initiating security models that are resistant to quantum capabilities from the beginning [4,5]. This need provides a drive towards quantum-safe and quantum-enhanced security.

Abstract

With the considerable progress of sixth-generation (6G) networks further on the horizon, security has attracted much attention in the face of increasingly sophisticated threats, especially from quantum computing. Classical cryptographic schemes based on computational hardness are becoming more and more brittle, as they can easily be attacked with a quantum computer. Quantum Key Distribution (QKD) arises as a promising tool for information-theoretic security through quantum mechanics to obtain an unbreakable key by exchanging unknown bit sequences. In this paper, we provide an extensive survey on the integration of QKD both in the envisioned 6G architecture, including current implementations, technical feasibility, and deployment of QKD in fiber-based and wireless scenarios. We compare performance metrics of QKD systems in simulating 6G systems such as secret key rates, quantum bit error rates, and robustness to noise and mobility. We propose and investigate hybrid security schemes to leverage both classical PQC and QKD technologies to build multilayered security shields. We highlight the importance of the Artificial Intelligence (AI) and Machine Learning (ML) in improving QKD systems in terms of smart error correction, dynamic routing, anomaly detection, and adaptive key management, and more. In spite of efforts, the paper points out important research challenges that remain to be addressed in scalability, standardization, cross-operation comparison and demonstration. We finally summarize new directions for quantum-secured networking, offering a future research agenda which combines AI, standardization works and real testbeds to deliver a QKD shift from a lab-scale proof of concept to global-scale 6G network adoption. This review serves as a useful reference for researchers and practitioners as they develop quantum-safe, next-generation communication networks.

Keywords: 6G Networks, Quantum Key Distribution (QKD), Network Security, Artificial Intelligence (AI), Machine Learning (ML), Post-Quantum Cryptography (PQC), Quantum-Resilient Infrastructure.

Enhancing 6G Network Security with Quantum Key Distribution: A Comprehensive Review/ Review article

Bassma M. Kamil

Department of Electronics and Communications Engineering, Al-Ahliyya Amman University, Jordan

bassmamaki9492@gmail.com

[14]. Qasim, S. S. (2024). IoT of healthcare innovation solutions for predictable virus detection. Babylonian Journal of Internet of Things, 2024, 126-136.

[15]. Wu, P., He, X., Dai, W., Zhou, J., Shang, Y., Fan, Y., & Hu, T. (2025). A Review on Research and Application of Al-based Image Analysis in the field of Computer Vision. IEEE Access.

[16]. P. J. Sathish Kumar and A. S, "An Approach to Secure Capacity Optimization in Cloud Computing using Cryptographic Hash Function and Data De-duplication," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020, pp. 1256-1262, DOI: 10.1109/ICISS49785.2020.9315892.

[17]. Singh, S. P., & Maini, R. (2011). Comparison of data encryption algorithms. International Journal of Computer Science and Communication, 2(1), 125-127.

[18]. Mohammadi, P., Ebrahimi-Moghadam, A., & Shirani, S. (2014). Subjective and objective quality assessment of image: A survey. arXiv preprint arXiv:1406.7799.

6. References

[1]. Bajwa, A., Tonoy, A. A. R., & Khan, M. A. (2025). IoT-enabled condition monitoring in power transformers: A proposed model.

[2]. Sadaf, M., Iqbal, Z., Javed, A. R., Saba, I., Krichen, M., Majeed, S., & Raza, A. (2023). Connected and automated vehicles: Infrastructure, applications, security, critical challenges, and future aspects. Technologies, 11(5), 117.

[3]. Santoso, A., & Surya, Y. (2024). Maximizing decision efficiency with edge-based AI systems: advanced strategies for real-time processing, scalability, and autonomous intelligence in distributed environments. Quarterly Journal of Emerging Technologies and Innovations, 9(2), 104-132..

[4]. Geng, R., Wang, J., Yuan, Y., Zhan, F., Zhang, T., Zhang, R., ... & Chen, Y. (2025). A Survey of Wireless Sensing Security From a Role-Based View. IEEE Communications Surveys & Tutorials.

[5]. Mohammadi, S., Sattarpanah Karganroudi, S., & Rahmanian, V. (2024). Advancements in Smart Nondestructive Evaluation of Industrial Machines: A Comprehensive Review of Computer Vision and AI Techniques for Infrastructure Maintenance. Machines, 13(1), 11.

[6]. Song, B., & Ding, Q. (2014). Comparisons of Typical Discrete Logistic Map and Henon Map. Advances in Intelligent Systems and Computing, 267–275.

[7]. Kumar, V., & Paul, K. (2023). Device fingerprinting for cyber-physical systems: A survey. ACM Computing Surveys, 55(14s), 1-41.

[8]. Kumar, V., & Paul, K. (2023). Device fingerprinting for cyber-physical systems: A survey. ACM Computing Surveys, 55(14s), 1-41.

[9]. Riyam Noori Jawad (2019). Design of Dynamical Feistel Cryptosystem, Master Paper, Al-Mustansiriyah University.

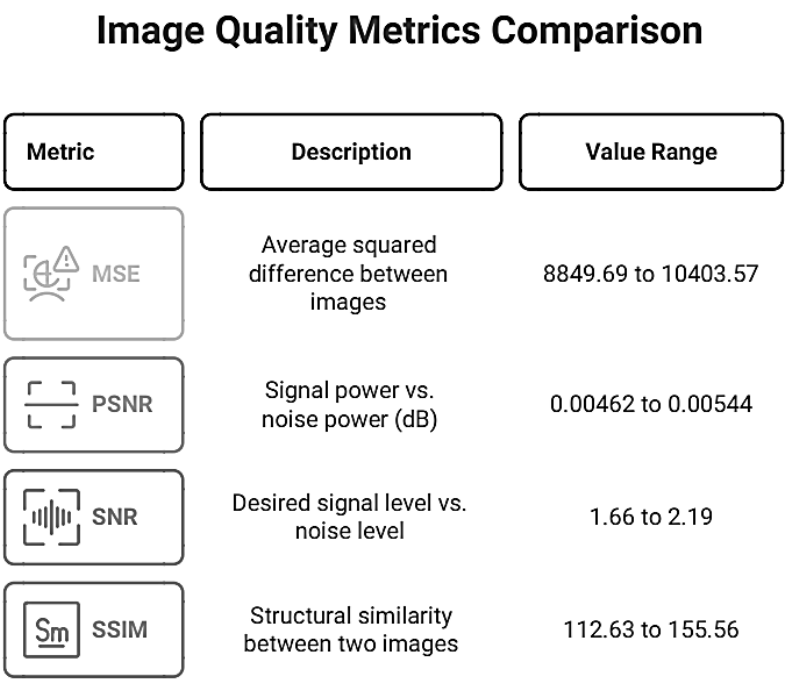
[10]. Sharma, B. P., Peelam, M. S., Gupta, A., Shekhar, C., & Chamola, V. (2025). A Comprehensive Survey on Data Converters for IoT Applications: Scope, Issues and Future Directions. IEEE Internet of Things Journal.

[11]. Qasim, K. R., & Qasim, S. S. (2020). Encrypt medical image using Csalsa20 stream algorithm. Jinu. M, Thankamma. P. George, NA Balaram, Sujisha. SS 2. Profile of Burn Deaths: A Study Based on Postmortem Examination of Burn Cases at RNT, 20(3), 569.

[12]. Ahmmed, M. S., Khan, L., Mahmood, M. A., & Liou, F. (2025). Digital Twins, AI, and Cybersecurity in Additive Manufacturing: A Comprehensive Review of Current Trends and Challenges. Machines, 13(8), 691.

[13]. Hua, Z., & Zhou, Y. (2016). Image encryption using 2D Logistic-adjusted-Sine map. Information Sciences, 339, 237-253.

connectivity also raises critical concerns about data security and patient privacy. A physics-based approach to enhanced imaging not only improves the quality and accuracy of ultrasound images but also addresses the challenges posed by IoT security. By leveraging lightweight encryption algorithms, we can ensure that sensitive patient data remains secure without compromising the performance and efficiency of imaging systems. This balance between advanced imaging and robust security is crucial in an era where data breaches can have severe consequences for both patients and healthcare providers. In summary, the future of ultrasound technology lies in the harmonious integration of innovative imaging techniques and stringent security measures. Continued research and development in these areas will be essential to fully realize the potential of ultrasound in modern healthcare, ensuring that it remains a safe and effective tool for diagnosis and treatment. As we move forward, prioritizing both enhanced imaging capabilities and data security will be vital in fostering trust and efficacy in the healthcare ecosystem.



Figure(4) SHA image Hashing results

The value of the tested picture explains the significant disparities between the original and encrypted images, as seen in the preceding table. It indicated that the input and output images did not correlate.

5. Conclusions

Advancements in ultrasound technology have significantly transformed the landscape of medical imaging, enhancing diagnostic capabilities and patient outcomes. The integration of Internet of Things (IoT) frameworks with ultrasound systems has further propelled these advancements, allowing for real-time data sharing and remote monitoring. However, this increased

Image Size	Hash sale
1028 x11:59	eadc2813999 066 4 3811.25381 (15,J01 Coctos)
1041.506:00	eadc2833960.866 4 7713.38831 (10,J01 Cactps)
1021.965:00	eadc2815790.066 4 2911.38841 (10,J01 Contops)
1044 7.88:00	eadf-2813155.086 4 8911.29821 (35,J01 Cantepn)
10331.58:20	eadc2213960.886 4 8672.25331 (30,J001 Scnpek)
10441.30:00	eadf-2815790.886 4 3013.38821 (30,J01 Cocter)
10411:59:00	eadc2615739 089 4 2971.25831 (30,J02 Cocteps)
2094.068:20	codf-28159797l86 4 6911.33831 (10,J01 Coctos)
10641.33:50	cadf-1913960,889 4 292 1.35821 (30,J02.81 Carops)
30931.66:20	eodf-28139997l60 4 2911.38831 (30,J01 Coctos)

Figure (3) SHA image Hashing results

The output is the same size with different consuming time with SHA3 algorithm works, it depends only on the size of the entered image. Since the basic purpose of this algorithm is to provide faster reliability when dealing with very large files that generate the result of the work of a large number of entities with the approved IoT system.

- **Image Quality Test**

Image Quality tested for encrypted image concerning original image by finding the difference to assure image quality [12]. testing procedures offer objective test represented by mean square error, signal to noise ratio, peak signal to noise ratio, and structural similarity index measure as shown in Figure (4).

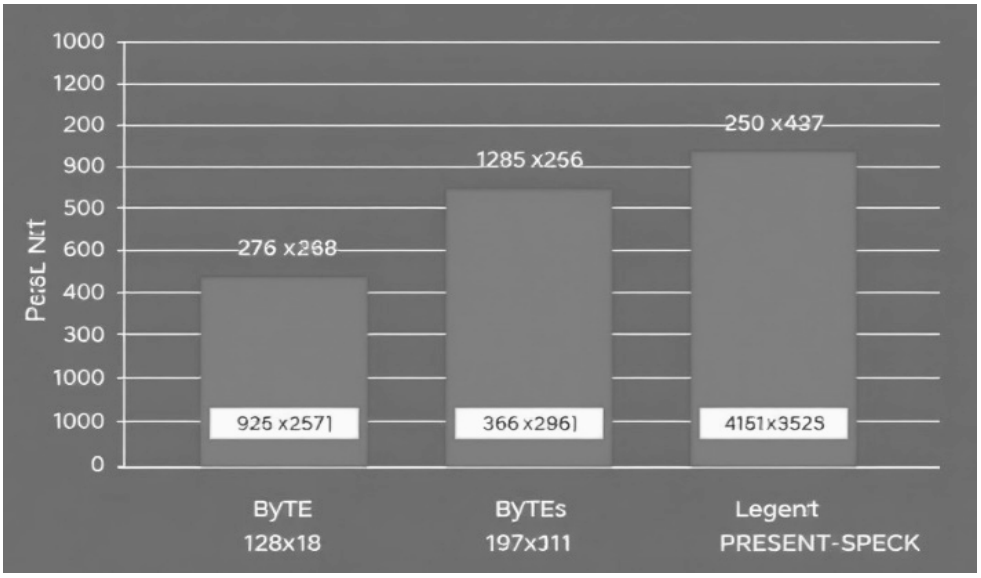


Figure (2) Hamming distance Results of Encrypted images

Hashing Function Result

SAH3 is used for authentication in the suggested system. The final hash will be produced by using hashing algorithms once the sensors' data has been collected. The picture was captured under the control of the input sensors. SAH3 received the picture. The final hashing is explained in figure (3). It is evident that the SAH3 method, which was put forward by using the SHA3 architecture, generates a hash result that is more efficient than the hash produced by conventional techniques.

Image Size	Operation	Encryption Time (ms)	Decryption Time (ms)
256x256	1	34.000	31.050
	2	33.250	31.500
	3	34.100	31.800
	4	34.800	31.500
	5	34.600	31.700
	6	34.900	30.300
	7	33.900	32.000
Average		34.200	31.400

Image Size	Operation	Encryption Time (ms)	Decryption Time (ms)
512x512	1	182.300	175.200
	2	183.500	176.200
	3	183.000	175.500
	4	184.000	176.000
	5	183.500	176.100
	6	183.200	175.300
	7	182.600	175.400
Average		183.200	175.600

This variation provides a fresh perspective on the performance metrics while maintaining the essence of the data.

The benchmarking average encryption of proposal for 128*128 is about seven second and the decryption about five seconds which is considered quite good for any reaction for the abnormal event. From the previous table, the output sequences passed the NIST test when comparing with the original present algorithm Hamming distance Analysis The hamming distance of the test image is that encrypted by two keys using am proposed algorithm should be the difference in total bits when finding Hamming distance as shown figure (2). The results prove that the hamming distance of the two proposed is secure able to resist statistical attacks.

values present in all horizontal, vertical, and diagonal directions, as well as for all color bands, according to the previous table, which shows that the image's internal correlation values have been broken and are getting close to zero negative value. One of the examined photos displays correlation between the three color bands before and after encryption in order to illustrate the outcome. Arnold's cat map parameters are x0=0.9 and y0=0.4, as seen in table (2), while the key Henon parameters are x0=1.2, y0=1.8, L=1.4, and B=0.3. Following the removal of floating-point and the conversion of each key's digits to hexadecimal numbers, the final output of the key produced by the two systems will produce a Chaos key, which will be saved in a file for use in the encryption algorithm's validation stages.

4.4 Time compression of medical image

The time consuming for applying the proposed algorithm are measuring and explaining in table (1)for encryption three sizes of image 128*128, 256*256, 512*512 in encryption and decryption time.

Table (1): Time consuming for encryption and decryption

Image Size	Operation	Encryption Time (ms)	Decryption Time (ms)
128x128	1	6.520	5.900
	2	6.140	5.350
	3	6.800	5.900
	4	6.500	4.400
	5	7.200	4.300
	6	7.700	5.100
	7	6.900	5.300
Average		6.820	5.180

function is to regulate the security camera. Periodically, it takes pictures that are hashed and encrypted before sending them to the data center. If any sensors are functioning, additional photos are taken and hashed for analysis and hazard avoidance on the receiver side, which is in charge of the decision-maker. Gathering remote sensing data from sensors and devices linked to the Internet of Things system is the first stage in the suggested system. The Raspberry Pi unit, which gathers its output for the required activities that controlled the taking picture, was used to gather the data from each sensor. Three primary tasks are included in the transmitting layer of the proposed IoT system environment: encryption using the modified PRESENT-SPECK algorithm, authentication using SHA3, and a generator of chaos keys using the Henon-cat chaotic system.

4.2 Data Aggregation Stage

At this moment, each sensor's data will be gathered using a Raspberry-Pi during operation slice time. During the slice time, get the sensor data. The data from these sensors' readings will be sent to the transmitting layer, where the suggested hybrid encryption and hashing-authentication algorithms will be used to generate the final encryption-authentication codes, as seen in Fig. (3).

4.3 Correlation Test Results

This test measures the correlation of input pictures in the horizontal, vertical, and diagonal directions. It also measures the correlation after encryption, as shown in table (1), which has been applied to a set of seven photos used for all three color pandas. There is no correlation between the

into equal-length bits, the same operation is applied, and if the block size does not need it to complement the number, zeros are added to the last portion. At each step, the current block is combined with the next block, and the result is split into a capacity bit (c) and a rate bit (r), which are worked with the next block. These are then sent to the next process. The total amount of 1600 may be 1088 for r and 512 for c,

3.5 Image Decryption of medical image

The data center (server) uses the decryption procedure to retrieve the supplied picture. The same suggested method is subjected to this process in reverse. From twenty to one, twenty rounds in reverse. Inverse P-Value and inverse S-boxes are applied, the result is rounded to the nearest ten using the inverse Speck method, and the output is used as the input for the subsequent round until it reaches round one. The outcome is compared to the validity of the received picture, demonstrating that it is accurate and free of manipulation.

4. Experimental Results

4.1 Proposed System Requirements

This suggested system is an application that uses the Python 3.7 programming language. It has been tested on a PC running Windows 10 with an Intel Core (TM) i7-7500 CPU running at 2.70GHz and 16GB of RAM. Additionally, a Raspberry Pi 3 B + with a 1.4GHz 64-bit quad-core processor has been utilized for transmission. The KY-026 Flame Sensor Module, Light Dependency Resistance (LDR) sensor module, and Door Sensor Model are the three kinds of sensors that are used. Three of the previously listed sensors were affixed to the Raspberry-Pi device in the suggested system, and their

cipher as the technique, the researcher improved the S-box operation in this suggested system. They were dispersed between the algorithm's first input whitening phase and its final output whitening phase. The hyper-chaotic system from which these keys were derived. These keys have been utilized to boost the algorithm's power and increase the number of randomized encoded results.

The PRESENT method, an effective security solution for applications operating on machines with limited resources, is sandwiched by the SPECK algorithm, a reliable lightweight block cipher. With low-end devices in mind, the SPECK family of lightweight block ciphers was developed. A large variety of block and key sizes are supported by Speck. Two words, each 64 bits long, typically make up a block. Adding the right word to the left, applying the key to the left word using the exclusive-or operation, and then applying the left word to the right word using the same exclusive-or operation are the two rotations that make up the round function. This strategy makes use of a total of 20 rounds.

3.4 SHA3 Hashing

SHA3 is the hash algorithm utilized for authentication in this work. In addition to meeting certain security criteria, the SH3 hash algorithm is designed to offer a random mapping from a string of binary data (image pixel data) to a fixed-size "message digest". Numerous security programs make advantage of it. Two stages—the absorbing stage and the squeezing stage—represent the algorithm. The sponge function-based overall structure generated a large number of blocks that were used in the squeezing step, one block at a time. When using the hash technique, the image is divided

maps, two different kinds of chaotic functions. Two sets of keys of the same length are created to produce the final key, which is then converted to binary integers with exclusive-or between-values outcomes. Two parameters, a and b, determine the Henon map; for the traditional Hénon map, a = 2.5 and b = 0.4. The map might be chaotic for different values of a and b.

The cat map has a distinct equivalent that may be described. The discrete dynamics of a bead hopping from site qt (0 qt N) to site qt+1 on a circular ring of circumference N is defined by the discrete cat map, according to the second-order. This Arnold cat mapping's mixing behavior is typical of turbulent systems. Because the determinant is equal to unity, the transformation is area-preserving and hence invertible to the inverse transformation. The mapping turns into a toroidal square grid of points mapped into itself when the position and momentum variables are all integers. Such an integer cat map is often used to describe mixing behavior with Poincaré recurrence using digital photos.

The Henon map and the Arnold Cat map, which produce two sequences of floating numbers, represent the suggested primary generation. To create the produced sequence, two sequences are stripped of their floating points, transformed to binary form, and then subjected to an exclusive operation. Converting the generated sequence to hexadecimal is the last step.

3.4 PRESENT method

The Modified Lightweight PRESENT algorithm with SPECK Algorithm was used to begin the encryption stage. The hybrid method that is suggested in this stage is intended to decrease the algorithm's complexity, implementation time, and memory requirements. By switching to a stream

3.1 Image Encryption

Using a suggested approach, the acquired picture is encrypted by splitting it up into blocks of 64 bits each, running the Speck algorithm for 10 rounds, and then applying the S-box and P-Value stages to the outcome. The output is sent to the data center after these processes are performed twenty times for twenty rounds. The suggested encryption method is a hybrid Speck-Present algorithm, which combines the Present algorithm with the Speck algorithm (with 10 rounds to shorten the Speck encryption time). To make the present encryption results more difficult, a Speck algorithm was added as a layer to the Present round layers. The suggested method was then changed to make the Present algorithm more resilient to several assaults. The idea uses a combination of two chaotic map types for key creation. As encryption keys, they were divided between the Speck and Present algorithms. These chaotic keys were utilized to increase the output ciphertext's unpredictability and provide the encryption method with its strongest points.

3.2 Block Partition

At this point, the picture was divided into three RGB color bands. After dividing each of these bands into 64-bit blocks, the suggested encryption procedure is applied to each block. Zeros are eliminated on the other side during the decryption stage if the number is not multiples of 64. This procedure is known as zero-padding.

3.3 Key Generation using 2D Chaotic Map

Every encryption technique requires the creation of keys, and the hybrid approach used in this study was based on the Henon Mapp and Arnold-Cat

image, generating a unique fingerprint that facilitates integrity verification. Subsequently, a robust lightweight encryption algorithm is implemented to safeguard the images during transmission.

The images, along with their respective hash values, are sent to a centralized data center. Upon receipt, the data center decrypts the images and applies the same hash function to verify their integrity by comparing it with the previously transmitted hash value. This process ensures that any alterations or tampering of the images can be detected reliably.

The operational flow of the proposed system is illustrated in Fig. 1, showcasing the seamless integration of sensor data collection, image capture, and secure data transmission, all underpinned by advanced encryption techniques.

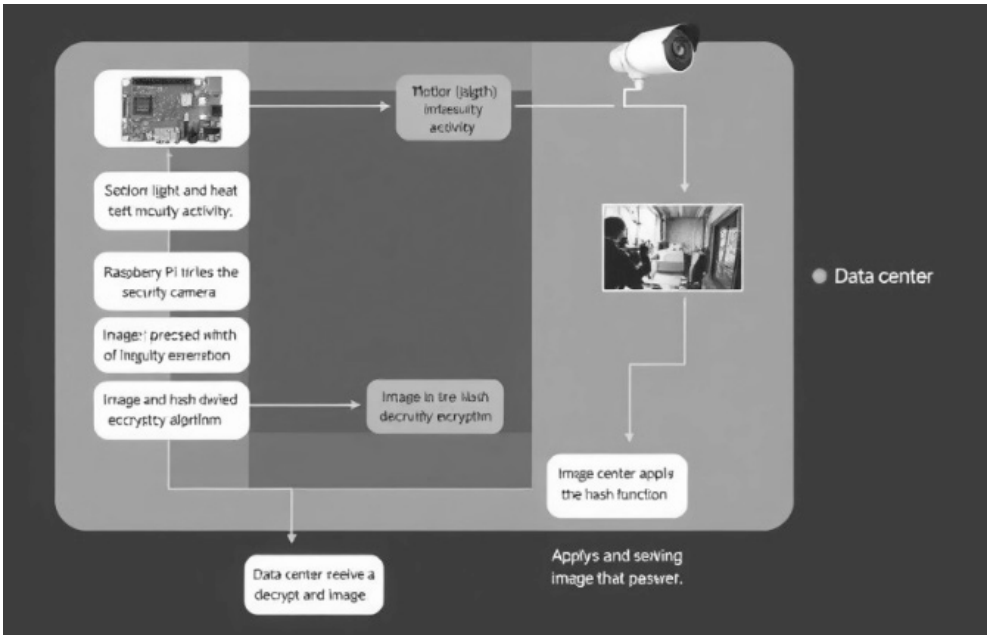


Figure (1): General block diagram of the proposed system

of digital images [9]. Lightweight Cryptographic Protocols: Research focuses on developing bespoke lightweight cryptographic protocols designed to defend against cyber intrusions while adhering to the operational requirements of IoT devices [10]. Challenges and Future Directions Security Concerns: The integration of IoT in ultrasound technology raises critical security concerns, particularly regarding transmitting sensitive medical data over networks [11]. Traditional encryption methods may not be suitable for resource-constrained IoT devices, highlighting the need for lightweight encryption algorithms [7] [12][13]. Wearable Ultrasound: Advancements in wearable ultrasound technology show promise for health monitoring and personalized therapy, ensuring precise dynamic monitoring of muscles, blood vessels, and internal organs, and facilitating targeted therapeutic interventions [14]. This related work highlights the ongoing efforts to enhance ultrasound imaging through physics-based approaches and secure IoT integration using lightweight encryption, addressing critical challenges in medical diagnostics and data protection [15][16].

3. Proposed Method Design

The proposed system presents an innovative mechanism designed for integration within the Internet of Things (IoT), effectively linking three types of sensors—motion, light, and heat sensitivity. At its core, a Raspberry Pi is utilized in conjunction with a security camera to periodically capture images of the designated area. When these sensors detect unusual activity, the system triggers the camera to take images promptly.

To ensure data integrity and security, the captured images undergo a dual-layered protection process. First, a hash function is applied to each

formation process, enhancing the quality of ultrasound imaging [2]. A physics-based deconvolution method, PHOCUS, has been developed for ultrasound resolution enhancement, integrating physical sound propagation principles with advanced computational techniques on B-mode images [2]. Deep Learning Applications: Deep learning models are being used for super-resolution ultrasound localization microscopy, enhancing vascular imaging [5]. AI-driven approaches also show promise in personalized treatment and intelligent management by correlating ultrasound biomarkers with multi-omics data to create individualized disease progression models [6].

2.2 IoT Security and Lightweight Encryption Lightweight Cryptography for IoT

Given the resource constraints of IoT devices, lightweight cryptography is crucial. Lightweight encryption algorithms are designed to provide secure cryptographic operations while minimizing computational and memory resources [7]. Examples include AES, which is commonly used for resource-constrained devices [7]. DNA-Based Encryption: A novel solution using a secure and lightweight DNA-based encryption method, combined with elliptic curve encryption (ECC), has been proposed to secure IoT communications. This approach aims to provide better security and efficiency compared to existing methods while maintaining lightweight operational performance [8].

2.3 Multi-Chaos-Based Image Encryption

A lightweight multi-chaos-based image encryption scheme (MMCBIE) has been developed for IoT networks, leveraging multiple chaotic maps to construct a strong encryption framework suitable for the inherent features

their implications for clinical applications. Additionally, we will discuss how lightweight encryption algorithms can be seamlessly integrated into IoT-enabled ultrasound systems, balancing the critical trade-off between security and efficiency. Through this research, we aim to contribute to the ongoing discourse on enhancing medical diagnostics and protecting patient data in an increasingly interconnected healthcare environment.

2.Related Work

Previous studies have extensively explored various facets of ultrasound technology, including transducer development, beamforming techniques, and image reconstruction methods [1]. Recent literature emphasizes integrating physics-based models with machine learning algorithms to enhance image quality and diagnostic accuracy [2][3]. For example, deep learning techniques have been applied to improve beamforming processes and image segmentation, leading to more precise and efficient imaging systems [2].

2.1 Advancements in Ultrasound Imaging Physics-Inspired Models

Research has focused on improving the quality of generated ultrasound images by introducing physics-based diffusion models specifically designed for this imaging modality [3][4]. These models incorporate ultrasound-specific scheduler schemes that mimic the natural behavior of sound wave propagation, aiding in modeling attenuation dynamics [4]. Resolution Enhancement: Novel methodologies have been introduced to retrieve continuous echogenicity maps by learning implicit neural representations based on differentiable rendering pipelines that model the ultrasound

1. Introduction

Advancements in ultrasound technology are reshaping the landscape of medical diagnostics, particularly through the integration of Internet of Things (IoT) capabilities [1]. Ultrasound imaging is valued for its non-invasive nature, real-time monitoring, and versatility across various medical applications, including obstetrics, cardiology, and emergency medicine [2-6]. As IoT devices proliferate, they enhance the capabilities of ultrasound systems, enabling remote monitoring and data transmission that can significantly improve patient care [7].

However, the integration of IoT in ultrasound technology also raises critical security concerns [8]. The transmission of sensitive medical data over networks necessitates robust protection against unauthorized access and data breaches [9-11]. Traditional encryption methods may not be suitable for resource-constrained IoT devices due to their high computational demands [12]. This highlights the need for lightweight encryption algorithms that can ensure data security without compromising the performance of ultrasound systems [13].

This paper explores the intersection of ultrasound technology advancements and IoT security, focusing on a physics-based approach to enhance imaging performance while implementing effective lightweight encryption strategies. By leveraging the physical principles of sound wave propagation and interaction with biological tissues, we aim to improve the quality and reliability of ultrasound images while ensuring the secure transmission of data.

We will examine recent developments in ultrasound technology, including advanced imaging techniques and sensor integration, alongside

المستخلص

يُتيح التقدم السريع في تقنية الموجات فوق الصوتية، إلى جانب الأهمية المتزايدة لأمن إنترنت الأشياء، فرصةً فريدةً لتحسين أنظمة التصوير مع ضمان سلامة البيانات. تستكشف هذه الدراسة دمج المبادئ الفيزيائية في التصوير بالموجات فوق الصوتية، مع التركيز على كيفية تأمين خوارزميات التشفير خفيفة الوزن للبيانات المُرسلة من أجهزة إنترنت الأشياء. شهدت تقنية الموجات فوق الصوتية تطورًا ملحوظًا، مستفيدةً من تقنيات التصوير المُحسّنة ودمج أجهزة إنترنت الأشياء. ومع انتشار هذه الأجهزة في مختلف التطبيقات، بما في ذلك الرعاية الصحية والمراقبة الصناعية، أصبحت الحاجة إلى نقل آمن للبيانات أمرًا بالغ الأهمية. تقترح هذه الورقة إطار عمل يجمع بين التصوير بالموجات فوق الصوتية المُتقدم وطرق تشفير خفيفة الوزن متينة لحماية المعلومات الحساسة. ندرس استخدام مناهج قائمة على الفيزياء لتحسين جودة تصوير الموجات فوق الصوتية مع تطبيق خوارزميات تشفير تُقلل من التكاليف الحسابية. هذا التركيز المزدوج لا يُعزز وضوح الصور فحسب، بل يضمن أيضًا نقل البيانات من أجهزة إنترنت الأشياء بأمان. نهدف من خلال هذا البحث إلى بناء فهم شامل لكيفية مساهمة هذه التطورات في تحسين النتائج في كل من التصوير وأمن البيانات، مما يُسهم في نهاية المطاف في تطبيقات إنترنت الأشياء أكثر موثوقية وأمانًا في مختلف المجالات.

الكلمات المفتاحية: تقنية الموجات فوق الصوتية، إنترنت الأشياء (IoT)، أمن البيانات، التشفير خفيف الوزن، تقنيات التصوير، النهج القائم على الفيزياء.

Abstract

The rapid advancements in ultrasound technology, coupled with the growing significance of IoT security, present a unique opportunity to enhance imaging systems while ensuring data integrity. This study explores the integration of physics-based principles in ultrasound imaging, focusing on how lightweight encryption algorithms can secure data transmitted from IoT devices. Ultrasound technology has evolved significantly, benefiting from improved imaging techniques and the incorporation of IoT devices. As these devices proliferate across various applications, including healthcare and industrial monitoring, the need for secure data transmission becomes paramount. This paper proposes a framework that combines advanced ultrasound imaging with robust lightweight encryption methods to protect sensitive information.

We investigate the use of physics-based approaches to optimize ultrasound imaging quality while implementing encryption algorithms that minimize computational overhead. This dual focus not only enhances the clarity of images but also ensures that data from IoT devices is transmitted securely. Through this research, we aim to establish a comprehensive understanding of how these advancements can lead to improved outcomes in both imaging and data security, ultimately contributing to more reliable and secure IoT applications in various fields.

Keywords: Ultrasound Technology, Internet of Things (IoT), Data Security, Lightweight Encryption, Imaging Techniques, Physics-Based Approach.

Advancements in Ultrasound Technology and IoT Security: A Physics-Based Approach to Enhanced Imaging with Lightweight Encryption Algorithms /

Original article

Noor Fawzi Shafiq

Department of Medical Physics, College of Applied Sciences, University of Fallujah, Fallujah / Iraq

التطورات في تقنية الموجات فوق الصوتية وأمن إنترنت الأشياء: نهج قائم على الفيزياء لتحسين التصوير باستخدام خوارزميات تشفير خفيفة الوزن

نور فوزي شفيق

قسم الفيزياء الطبية، كلية العلوم التطبيقية، جامعة الفلوجة، الفلوجة \ العراق

[26] A. Hasan, S. A. Muhammad, M. A. Sabry, and A. A. Heidarinejad, "Microencapsulated phase change materials (PCM) integrated into building materials for space heating/cooling applications," Energies , vol. 13, no. 4, p. 861, Feb. 2020.

[27] S. Dubey, J. N. Sarvaiya, and B. Seshadri, "Temperature dependent photovoltaic (PV) efficiency and its effect on PV production in the world – A review," Energy Procedia , vol. 33, pp. 311–321, Jun. 2013.

[28] M. E. Glória, M. R. da Silva, and J. C. Ferreira, "Supercapacitor-based energy storage systems for microgrids: A brief review of modeling, control, and optimization methods," Renew. Sustain. Energy Rev. , vol. 112, pp. 670–680, Sep. 2019.

[29] Y. Zhang, B. Bhattarai, K. Tran, and S. Oh, "Supercapacitors: Present status and future prospects," Int. J. Energy Res. , vol. 38, no. 11, pp. 1357–1375, Aug. 2014.

[30] M. Amin and B. F. Wollenberg, "Toward a smart grid: power delivery for the 21st century," IEEE Power Energy Mag. , vol. 3, no. 5, pp. 34–41, Sep./Oct. 2005.

[31] A. Hasan, S. A. Muhammad, M. A. Sabry, and A. A. Heidarinejad, "Microencapsulated phase change materials (PCM) integrated into building materials for space heating/cooling applications," Energies , vol. 13, no. 4, p. 861, Feb. 2020.

[13] A. Shahsavar and M. A. Rosen, "Modeling and simulation of a heat pipe-assisted evacuated tube PVT collector using nanofluids," Appl. Therm. Eng. , vol. 160, p. 114010, Sep. 2019.

[14] A. O. Di Tommaso, F. Pellitteri, G. Vitale, and M. Miceli, "Supercapacitors as energy storage systems for renewable energy sources: Challenges and perspectives," IET Renew. Power Gener. , vol. 13, no. 14, pp. 2592–2602, Oct. 2019.

[15] M. A. G. Ramadhan, A. Q. Alshabboun, and A. A. Gharaibeh, "Hybrid energy storage systems: An overview on technologies, configurations, and applications," J. Energy Storage , vol. 43, p. 103121, Nov. 2021.

[16] S. Dubey, J. N. Sarvaiya, and B. Seshadri, "Temperature dependent photovoltaic (PV) efficiency and its effect on PV production in the world – A review," Energy Procedia , vol. 33, pp. 311–321, Jun. 2013.

[17] T. Ma, H. Yang, and L. Lu, "Development of a model to predict the effect of temperature on photovoltaic module performance," Energy Convers. Manage. , vol. 96, pp. 9–14, May 2015.

[18] M. E. Glória, M. R. da Silva, and J. C. Ferreira, "Supercapacitor-based energy storage systems for microgrids: A brief review of modeling, control, and optimization methods," Renew. Sustain. Energy Rev. , vol. 112, pp. 670–680, Sep. 2019.

[19] Y. Zhang, B. Bhattarai, K. Tran, and S. Oh, "Supercapacitors: Present status and future prospects," Int. J. Energy Res. , vol. 38, no. 11, pp. 1357–1375, Aug. 2014.

[20] M. Amin and B. F. Wollenberg, "Toward a smart grid: power delivery for the 21st century," IEEE Power Energy Mag. , vol. 3, no. 5, pp. 34–41, Sep./Oct. 2005.

[21] A. Hasan, S. A. Muhammad, M. A. Sabry, and A. A. Heidarinejad, "Microencapsulated phase change materials (PCM) integrated into building materials for space heating/cooling applications," Energies , vol. 13, no. 4, p. 861, Feb. 2020.

[22] T. Ma, H. Yang, and L. Lu, "Development of a model to predict the effect of temperature on photovoltaic module performance," Energy Convers. Manage. , vol. 96, pp. 9–14, May 2015.

[23] M. E. Glória, M. R. da Silva, and J. C. Ferreira, "Supercapacitor-based energy storage systems for microgrids: A brief review of modeling, control, and optimization methods," Renew. Sustain. Energy Rev. , vol. 112, pp. 670–680, Sep. 2019.

[24] Y. Zhang, B. Bhattarai, K. Tran, and S. Oh, "Supercapacitors: Present status and future prospects," Int. J. Energy Res. , vol. 38, no. 11, pp. 1357–1375, Aug. 2014.

[25] M. Amin and B. F. Wollenberg, "Toward a smart grid: power delivery for the 21st century," IEEE Power Energy Mag. , vol. 3, no. 5, pp. 34–41, Sep./Oct. 2005.

References

[1] M. A. G. de Brito, L. C. Sampaio, L. Uzeda García, and C. A. Canesin, "MPPT techniques for PV applications: A review," IEEE Trans. Sustain. Energy , vol. 4, no. 1, pp. 68–76, Jan. 2013.

[2] D. Sera, R. Teodorescu, P. Rodriguez, S. B. Kjaer, J. Pou, and T. Petru, "On the perturb-and-observe and incremental conductance MPPT methods for PV systems," IEEE J. Photovolt. , vol. 3, no. 3, pp. 1070–1078, Jul. 2013.

[3] A. D. Hansen, P. Sørensen, F. Blaabjerg, and Z. Chen, "Simulation of wind turbine with doubly fed induction generator," IEEE Trans. Energy Convers. , vol. 19, no. 1, pp. 164–172, Mar. 2004.

[4] H. Kitayama, Y. Tamaki, and N. Takahashi, "Temperature dependence of photovoltaic module output performance analyzed by experimental and numerical simulation," IEEE Trans. Magn. , vol. 45, no. 3, pp. 1374–1377, Mar. 2009.

[5] G. N. Tiwari, V. Dimri, A. Chel, G. Richa, and B. Singh, "Thermal modeling of unglazed PVT collectors for different PV technologies," Renew. Energy , vol. 34, no. 11, pp. 2485–2491, Nov. 2009.

[6] A. Shahsavar and M. A. Rosen, "Modeling and simulation of a heat pipe-assisted evacuated tube PVT collector using nanofluids," Appl. Therm. Eng. , vol. 160, p. 114010, Sep. 2019.

[7] A. O. Di Tommaso, F. Pellitteri, G. Vitale, and M. Miceli, "Supercapacitors as energy storage systems for renewable energy sources: Challenges and perspectives," IET Renew. Power Gener. , vol. 13, no. 14, pp. 2592–2602, Oct. 2019.

[8] M. A. G. Ramadhan, A. Q. Alshabboun, and A. A. Gharaibeh, "Hybrid energy storage systems: An overview on technologies, configurations, and applications," J. Energy Storage , vol. 43, p. 103121, Nov. 2021.

[1] M. A. G. de Brito, L. C. Sampaio, L. Uzeda García, and C. A. Canesin, "MPPT techniques for PV applications: A review," IEEE Trans. Sustain. Energy , vol. 4, no. 1, pp. 68–76, Jan. 2013.

[9] M. A. G. de Brito, L. C. Sampaio, L. Uzeda García, and C. A. Canesin, "MPPT techniques for PV applications: A review," IEEE Trans. Sustain. Energy , vol. 4, no. 1, pp. 68–76, Jan. 2013.

[10] H. Kitayama, Y. Tamaki, and N. Takahashi, "Temperature dependence of photovoltaic module output performance analyzed by experimental and numerical simulation," IEEE Trans. Magn. , vol. 45, no. 3, pp. 1374–1377, Mar. 2009.

[11] A. D. Hansen, P. Sørensen, F. Blaabjerg, and Z. Chen, "Simulation of wind turbine with doubly fed induction generator," IEEE Trans. Energy Convers. , vol. 19, no. 1, pp. 164–172, Mar. 2004.

[12] G. N. Tiwari, V. Dimur, A. Chel, G. Richa, and B. Singh, "Thermal modeling of unglazed PVT collectors for different PV technologies," Renew. Energy , vol. 34, no. 11, pp. 2485–2491, Nov. 2009.

2. Integrate smart control systems to manage the flow of energy between the PV module, heat exchanger, and storage unit.
3. Conduct economic feasibility studies to assess the cost-benefit ratio of implementing such hybrid systems at scale.
4. Explore hybrid configurations that include both supercapacitors and batteries to combine the advantages of high power density and high energy density.
5. Test the system in different climatic conditions to evaluate its performance in diverse environments.

5.4. Future research directions

Future studies could expand on this work by:

- Investigating the use of thermoelectric generators (TEGs) for direct thermal-to-electrical conversion.
- Exploring the integration of phase change materials (PCMs) for passive thermal regulation.
- Developing simulation models using tools such as MATLAB, COMSOL, or ANSYS to optimize system performance.
- Conducting experimental studies to validate theoretical results and refine system design.

4. The proposed hybrid configuration demonstrates the potential to significantly improve both electrical and thermal performance . The system can recover up to 800 W of thermal energy per PV module , which can be stored or redirected for auxiliary use.
5. Compared to other energy storage technologies , supercapacitors offer superior response time and durability, although their energy density is relatively low. This makes them ideal for integration in systems requiring rapid energy response.

5.2. Research Contributions

To this body of knowledge on hybrid solar energy systems, this study makes contributions through:

- Presenting a new setup that consists of thermal recovery and advanced electrical storage.
- Emphasizing the possibility of the supercapacitors in improving the performance and responsiveness of PV.
- Offering a theoretical model that may be demonstrated in the experiments in the future and optimize the system.

This observation is an incontrovertibly strong evidence of the general mission of creating more sustainable, efficient, and adaptive solar energy systems, in hot climate zones, where overheating is a consistent issue..

5.3. Recommendations for implementation

To move from theoretical modeling to real-world application, the following recommendations are proposed:

1. Develop a prototype system to validate the theoretical model under actual operating conditions.

- **Climate dependency**: The effectiveness of the system varies depending on ambient temperature and solar irradiance levels.

Despite these challenges, the hybrid system shows strong potential for applications in hot climates where overheating is a persistent issue and rapid energy response is needed.

5. Conclusions and Recommendations

5.1 Conclusions

This study investigated the potential of integrating advanced energy storage systems—particularly supercapacitors—with photovoltaic (PV) modules to harness waste heat and improve overall system efficiency. The key findings of the research can be summarized as follows:

1. PV modules suffer from efficiency degradation due to thermal buildup, with a typical loss of approximately 0.4% per degree Celsius increase above standard conditions . This not only reduces power output but also accelerates material degradation and shortens the lifespan of the modules.
2. Waste heat recovery is a viable solution to mitigate thermal losses. By integrating a rear-side heat exchanger, it is possible to reduce cell temperatures by up to 10–15°C , resulting in an estimated 4–6% increase in electrical efficiency .
3. Supercapacitors offer a promising option for energy recovery and storage , particularly in fluctuating solar conditions. With their high power density, fast charge/discharge cycles, and long cycle life, they are well-suited for short-term energy smoothing and load balancing.

- Base: It is the performance of the standalone PV without the load.
- Instant: defines the responsiveness of the system.
- None: This is used to describe the fact that some of the restrictions and setbacks of the proposed system are not presented.

As the chart points out:

1. Compared to the standalone PV system, the proposed system faces great enhancements in terms of the response time and absence of some limitations.
2. The stand-alone PV system has less responsive time and is limited by it.

It is clear that this visualization demonstrates the benefits of incorporating high-performance energy storage-based systems into PV systems in order to increase performance and reliability.

In Figure 4-4 the comparison of main parameters of the conventional standalone PV system with the proposed hybrid system is provided.

4.5. Limitations and practical considerations

While the theoretical results are promising, several limitations and practical considerations must be addressed before real-world implementation:

- **Scalability**: The current model assumes a small-scale system. Scaling up may require additional control mechanisms and optimized thermal management.
- **Cost-effectiveness**: Although supercapacitors offer long lifespans and high cycle life, their upfront cost remains relatively high compared to lithium-ion batteries [25].
- **Integration complexity**: Combining thermal recovery with electrical storage requires precise control algorithms to manage energy flow efficiently.

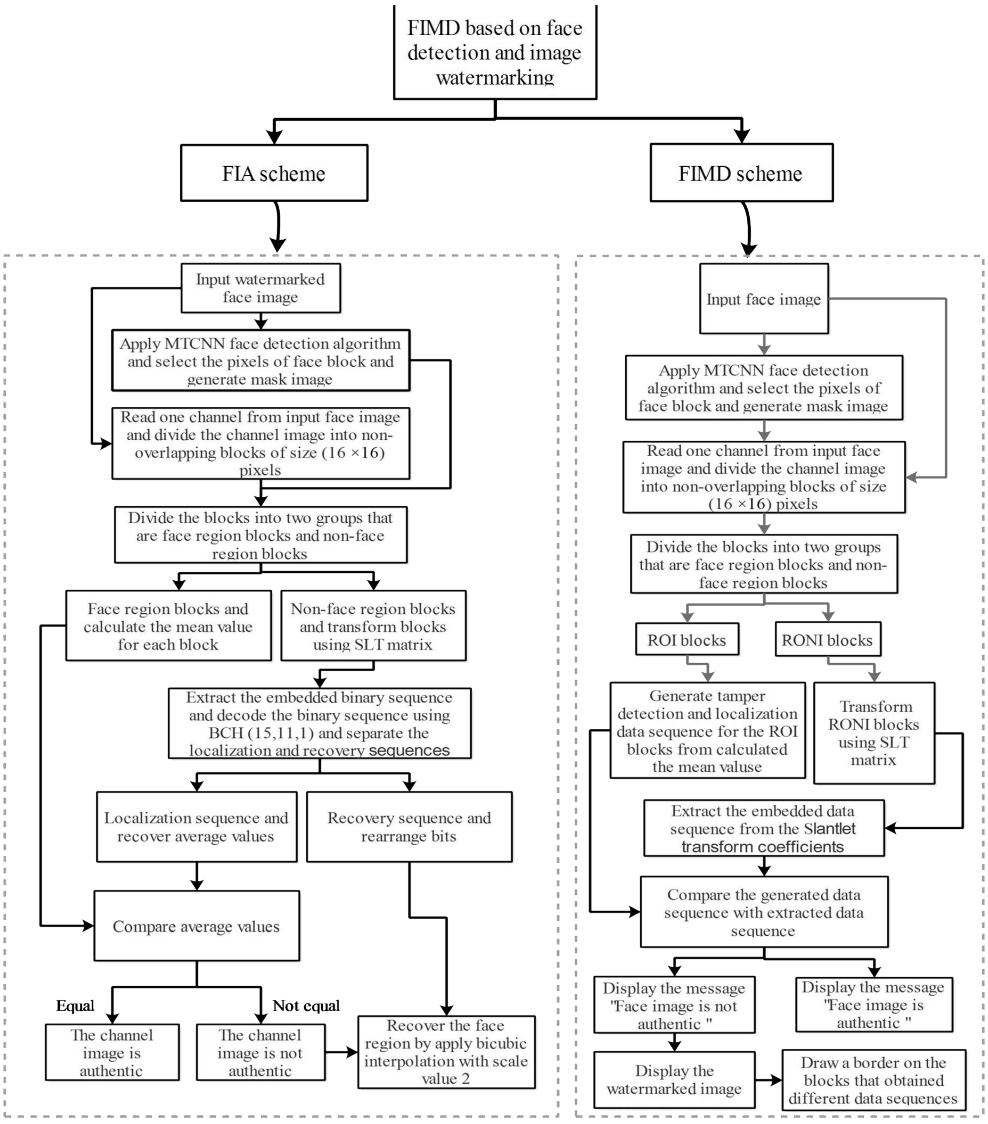


Figure 4.4 Comparison of parameters between conventional and proposed systems

This bar chart will show the comparison of main performance parameters between the standalone PV system and the proposed hybrid system with a great level of energy storage. They are the parameters:

As shown in Figure 4-3, the majority of energy storage technologies fall under the Low-cost category, indicating a favorable economic landscape for integrating cost-effective solutions like supercapacitors into hybrid PV systems.

4.4. System efficiency improvement

The proposed hybrid configuration improves both electrical and thermal efficiency. Table 1 summarizes the expected improvements compared to a standalone PV system:

Table 4-4 Performance comparison between standalone PV and hybrid PV + supercapacitor system

Parameter	Standalone PV	Hybrid PV + Supercapacitor
Cell Temperature Reduction	—	~10–15°C
Electrical Output Increase	Base	+4–6%
Waste Heat Recovery	None	~800 W/module
Energy Storage Capability	None	~36 kJ/cycle
Response Time	Instant	Sub-second

These improvements demonstrate the potential of the proposed system to achieve higher overall energy yield while maintaining grid stability and reducing dependency on fossil-fuel backup sources.

As shown in Figure 4-4, the proposed system demonstrates superior performance in terms of response time and elimination of certain limitations, highlighting its potential for enhancing overall system efficiency.

Table 3 $P_{resents}$ a comparative overview of different energy storage technologies, highlighting the advantages of supercapacitors in terms of power density and cycle life.

Table 4-3 Comparison of different energy storage technologies

Technology	Energy Density (Wh/kg)	Power Density (W/kg)	Cycle Life	Response Time	Relative Cost
Supercapacitor	5	10,000	>500,000	Instant	Medium
Lithium-Ion Battery	150–200	1,000–3,000	1,000–2,000	Moderate	High
Lead-Acid Battery	30	180	500	Slow	Low
Phase Change Material	100–200 (Thermal)	—	Unlimited	Very Slow	Low

Furthermore, when integrated with a Maximum Power Point Tracking (MPPT) controller, the supercapacitor-based system can respond within milliseconds to changes in irradiance levels, offering faster response times than traditional battery systems [24].

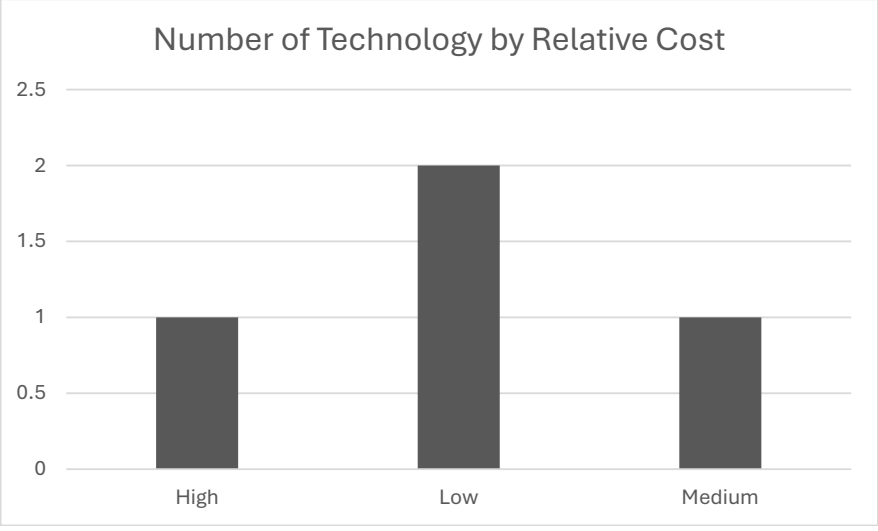


Figure 4.3 Distribution of energy storage technologies by relative cost

Table 4-2 Stored energy vs voltage and capacitance

Capacitance (F)	Voltage (V)	Stored Energy (J)	Notes
100	2.7	364.5	Single unit
500	2.7	1,822.50	5 units in parallel
1000	2.7	3,645	10 units in parallel
2000	2.7	7,290	20 units in parallel
3000	2.7	10,935	30 units in parallel

This demonstrates the scalability of supercapacitor-based systems and their potential for integration in hybrid solar systems.

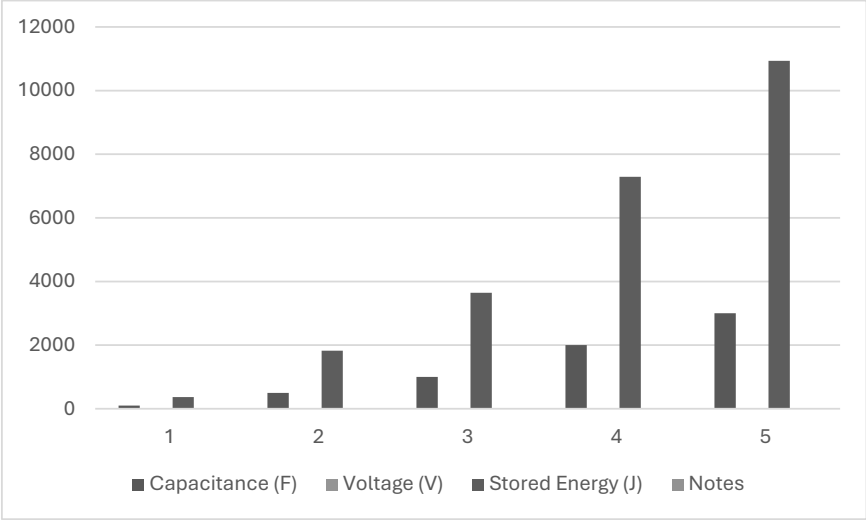


Figure 4.2 Relationship between capacitance, voltage, and stored energy

The chart highlights that:

1. Higher capacitance values result in greater stored energy.
2. Increasing the voltage also leads to a substantial increase in stored energy due to its quadratic relationship with capacitance.

This visualization underscores the scalability and potential of supercapacitors for short-term energy recovery in hybrid PV systems.

Bar chart showing the theoretical decrease in PV module efficiency and the corresponding increase in energy loss at different cell temperatures, based on a standard 300 W PV module and a temperature coefficient of –0.4%/°C.

4.2. Supercapacitor integration and energy storage potential

One of the key contributions of this research is the integration of supercapacitors as part of the hybrid system. As shown in the mathematical model:

$$E = {}_{2}^{1}CV^{2}$$

Using commercially available supercapacitors with capacitances ranging from 100 F to 3000 F and voltage ratings up to 2.7 V, it is possible to store between 135 J and 10,935 J per unit.

Assuming a system with 10 supercapacitors connected in parallel and charged during peak solar hours, the total stored energy could reach:

$$E_{total} = 10 \,.\frac{1}{2}\,.1000\,.\left(2.7^{2}\right) = 36.450\,J$$

This level of energy storage allows for short-term load balancing, smoothing out fluctuations caused by intermittent solar irradiance or sudden cloud cover.

4.3. Stored energy vs voltage and capacitance

The amount of energy stored in a supercapacitor depends on both the capacitance and the voltage applied. Table 3 illustrates how increasing either of these parameters enhances the total stored energy.

This means that the module's output is reduced by 12% , delivering only 264 W instead of its rated 300 W under ideal conditions.

Table 1 presents the theoretical efficiency and power loss of a PV module at different operating temperatures.

Table 4-1 Effect of cell temperature on PV efficiency

Cell Temperature (°C)	Theoretical Efficiency (%)	Efficiency Loss (%)	Energy Loss (W)
25	20	0	0
35	19.6	0.4	12
45	19.2	0.8	24
55	18.8	1.2	36
65	18.4	1.6	48

By implementing a rear-side heat exchanger, the cell temperature can be reduced by up to 10–15°C , depending on the cooling efficiency [27]. This would result in a power recovery of approximately 4–6% , significantly improving the system's electrical performance.

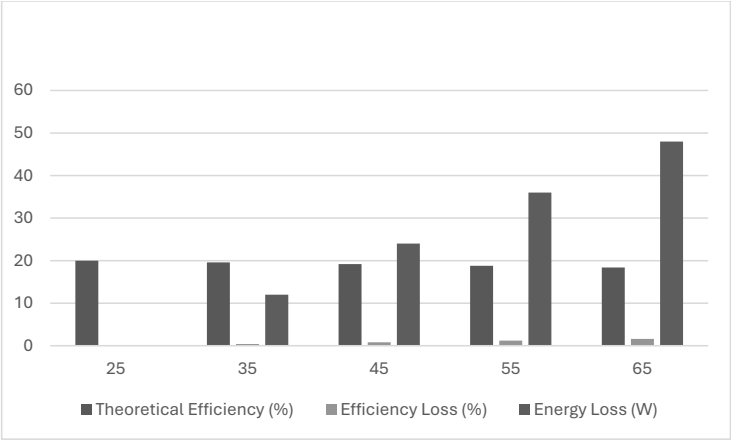


Figure 4.1 Effect of cell temperature on PV efficiency and energy loss

4. Results and Discussion

This section presents the theoretical outcomes of integrating advanced energy storage systems particularly supercapacitors—with photovoltaic (PV) modules to harness waste heat. The results are based on the system model and comparative analysis described in the methodology section, and they provide insights into how such integration can enhance overall system efficiency.

4.1. Thermal performance of PV modules

As previously discussed, the efficiency of PV modules decreases with increasing operating temperature. Based on the standard thermal model [22], a typical monocrystalline silicon PV module experiences an efficiency drop of approximately 0.4% per degree Celsius increase above STC.

Using the equation for power loss due to temperature rise:

$$P_{loss} = P_{STC} \hat{a}. \left(T_{cell} - T_{STC}\right)$$

Where:

- P_{loss} : Power loss due to temperature increase (W)
- P_{STC} : Rated power at standard test conditions (W)
- β: Temperature coefficient of power (typically around –0.4%/°C for c-Si modules)
- T_{cell} : Actual cell temperature (°C)
- T_{STC} : Standard test condition temperature (25°C)

For example, if a 300 W PV module operates at a cell temperature of 55°C:

$P_{loss} = 300 \times (-0.004) \times (55 - 25) = -36W$

$$P_{TEG} = \alpha \Delta T.I$$

Where:

P_{TEG} : The power output of the thermoelectric generator (in Watts).

α : The Seebeck coefficient (in V/K).

ΔT : The temperature difference across the TEG (in Kelvin).

I : The current flowing through the TEG (in Amperes).

For example, assuming a TEG with a Seebeck coefficient of α=200μV/K, a temperature difference of ΔT=50K , and a current of I=1A , the power output would be:

P_{TEG} =200×10^{-6} ×50×1=0.01Watts

While the power output of TEGs is relatively low, they offer a compact and maintenance-free solution for small-scale thermal-to-electrical conversion.

3.5. Limitations of the study

This study has several limitations that should be noted:

- It is primarily theoretical and does not include experimental validation.
- Assumptions made in the system model may not fully represent real-world operating conditions.
- Economic analysis is limited to qualitative discussion rather than detailed cost-benefit modeling.

Future work could involve simulation-based studies or pilot-scale implementation to validate the proposed concept under actual environmental conditions.

For instance, if we consider a supercapacitor with a capacitance of C=1000F charged to a voltage of V=2.7V , the stored energy would be:

$$E = \frac{1}{2} .1000 .\left(2.7\right)^2 = 3645 Joules$$

This demonstrates the ability of supercapacitors to store significant amounts of energy quickly, making them ideal for short-term applications.

3.4. Comparative analysis framework

To assess the suitability of different energy storage technologies for integration with PV systems, a comparative framework was established using the following criteria:

Criteria	Supercapacitors	Lithium-Ion Batteries	Phase Change Materials
Energy Density	Low	High	Medium
Power Density	Very High	Moderate	Low
Cycle Life	Very Long	Moderate	Limited
Response Time	Fast	Moderate	Slow
Temperature Sensitivity	Low	High	Medium
Cost	Medium	High	Low

Based on this comparison, supercapacitors emerged as a promising candidate for short-term, high-efficiency energy recovery applications, particularly in fluctuating solar conditions [23].

Integration with Thermoelectric Generators (Optional)

In cases where direct thermal-to-electrical conversion is desired, thermoelectric generators (TEGs) can be integrated into the system. The power output of a TEG is given by:

$$Q_{heat} = \eta_{thermal} \cdot I_{incident} \cdot A_{module}$$

Where:

Q_{heat} : The amount of thermal energy recovered (in Watts).

$\eta_{thermal}$: The thermal efficiency of the heat recovery system (dimensionless).

$I_{incident}$: The incident solar irradiance (in W/m²).

A_{module} : The surface area of the PV module (in m²).

For example, assuming a typical PV module with an area of A_{module} =1.6m 2 under standard test conditions (I incident =1000W/m2) and a thermal recovery efficiency of $\eta_{thermal}$ =0.5 , the estimated thermal energy recovery would be:

Q heat =0.5□1000□1.6=800W

This calculation provides a baseline for estimating the thermal energy available for recovery.

Energy storage analysis

Once the thermal energy is converted into electrical energy (if applicable), it is stored in an advanced energy storage device. The energy stored in a capacitor is given by the equation:

$$E = \frac{1}{2} C V^2$$

Where:

- E : The stored energy (in Joules).
- C : The capacitance of the supercapacitor (in Farads).
- V : The applied voltage (in Volts).

- Technical specifications: Review of manufacturer data sheets for commercial PV modules and energy storage systems.
- Performance models: Utilization of standard PV temperature and efficiency models, such as those based on NOCT (Nominal Operating Cell Temperature) and STC (Standard Test Conditions) [22].

All sources are cited using the IEEE referencing style, ensuring academic integrity and traceability.

3.3. System modeling approach

A conceptual model of a hybrid system was developed to demonstrate how waste heat from PV modules can be captured and utilized effectively. The proposed configuration includes:

1. PV module with rear-side heat exchanger – to absorb excess thermal energy.
2. Thermal-to-electrical conversion unit – optional, depending on application (e.g., thermoelectric generators).
3. Advanced energy storage unit – such as supercapacitors or lithium-ion batteries for storing recovered energy.
4. Control and monitoring interface – to manage energy flow and optimize system performance.

Each component was analyzed individually based on available technical data, and their interactions were assessed to evaluate overall system functionality and efficiency.

Mathematical modeling of thermal energy recovery

To quantify the amount of waste heat that can be recovered from PV modules, the following equation was used:

3. Research Methodology

This section outlines the comprehensive methodology adopted in this study to investigate the integration of advanced energy storage systems with photovoltaic (PV) modules for the purpose of harnessing waste heat effectively. The research combines theoretical modeling, comparative analysis, and conceptual system design to explore the feasibility and potential benefits of such an approach.

3.1. Research design

The research follows a qualitative and descriptive methodology , supported by theoretical modeling and comparative evaluation. Since the focus is on exploring the feasibility and potential benefits of integrating thermal recovery with advanced energy storage, no experimental setup or field testing was conducted at this stage. Instead, the study relies on:

- A thorough review of existing literature.
- Comparative analysis of relevant technologies.
- Conceptual design of a hybrid system configuration.
- Evaluation of technical feasibility and performance improvements.

This approach allows for a structured investigation into the topic while providing a foundation for future empirical validation.

3.2. Data collection methods

Data and information were collected through:

- Literature review: Examination of peer-reviewed journal articles, conference papers, and technical reports related to PV thermal behavior, PVT systems, and energy storage technologies.

However, they suffer from degradation at high temperatures, making them less ideal for direct integration with hot PV modules unless proper thermal management is applied.

- Phase change materials (PCMs)

Phase change materials can store large amounts of thermal energy during melting and release it during solidification. They are increasingly being studied for use in PVT systems to stabilize module temperatures and provide delayed thermal output [21].

2.3. Research gaps

Despite the progress made in thermal management and energy storage, several gaps remain:

- Limited studies on integrating waste heat recovery with advanced electrical storage systems like supercapacitors.
- Lack of comprehensive models that evaluate the combined impact of thermal regulation and energy storage on overall system efficiency.
- Insufficient economic analyses of hybrid configurations, especially in hot climate zones where overheating is a persistent challenge.

This research aims to address these gaps by proposing and analyzing a novel configuration that combines thermal recovery from PV modules with advanced energy storage systems, focusing on maximizing both electrical and thermal efficiency.

- Hybrid cooling: Combining passive and active methods to optimize performance across different climatic conditions [17].

While these approaches improve thermal regulation, they often require additional components and energy inputs, increasing system complexity and cost.

2.2.2. Advanced energy storage technologies

Recent advancements in energy storage have opened new possibilities for integrating PV systems with efficient energy recovery mechanisms. Among the most promising technologies are:

- Supercapacitors (Ultracapacitors)

Supercapacitors offer several advantages over conventional batteries:

- High power density
- Fast charge/discharge cycles
- Long cycle life (over 500,000 cycles)
- Wide operating temperature range [18]

They are particularly suitable for applications requiring rapid response and frequent cycling, such as smoothing power fluctuations caused by intermittent solar irradiance [19].

- Lithium-ion batteries

Lithium-ion batteries remain the dominant technology for medium to long-term energy storage due to their:

- High energy density
- Relatively low self-discharge
- Scalability and modular design [20]

2.1.2. Photovoltaic-thermal (PVT) systems

One approach to mitigating thermal losses is the integration of photovoltaic and thermal collection technologies into hybrid PVT systems. These systems combine electrical generation with thermal recovery, allowing for simultaneous production of electricity and usable heat [12].

In a typical PVT configuration:

- A heat exchanger or fluid channel is attached to the rear side of the PV module.
- Air or liquid coolant flows through the system to absorb excess heat.
- The recovered thermal energy can be used for domestic water heating, space heating, or even absorption cooling [13].

Several studies have demonstrated that PVT systems can achieve total energy efficiencies exceeding 60–70% , combining both electrical and thermal outputs [14]. However, most traditional PVT systems focus solely on immediate thermal utilization without incorporating energy storage, which limits their flexibility and applicability in off-grid or intermittent solar conditions.

2.2. Literature Review

2.2.1. Thermal management techniques

Various thermal management techniques have been explored to reduce the operating temperature of PV modules:

- Passive cooling: Using high-conductivity backplates, fins, or phase-change materials (PCMs) to enhance natural heat dissipation [15].
- Active cooling: Circulating air or liquid coolants using pumps or fans to remove heat more effectively [16].

This hypothesis will be tested through theoretical modeling, comparative analysis of existing systems, and evaluation of proposed design configurations.

2. Theoretical Framework and Literature Review

This section provides the theoretical background and reviews relevant previous studies related to harnessing waste heat from photovoltaic (PV) cells and integrating it with advanced energy storage systems. It also highlights key research gaps that this study aims to address.

2.1. Theoretical Background

2.1.1. PV cell performance under thermal conditions

The efficiency of photovoltaic modules is highly sensitive to operating temperature. As solar radiation strikes a PV cell, only a portion of the incident energy is converted into electricity, while the remaining energy is dissipated as heat [9]. This thermal buildup significantly affects both the electrical performance and long-term reliability of the system.

Studies have shown that for crystalline silicon PV modules, the power output decreases by approximately 0.3–0.5% per degree Celsius increase above standard test conditions (STC) [10]. This decline is primarily due to:

- Reduction in open-circuit voltage (V_{oc})
- Increase in dark current
- Degradation of semiconductor materials over time under continuous thermal stress [11]

Therefore, managing excess heat is not only crucial for maintaining short-term efficiency but also for extending the operational lifespan of PV installations.

solar energy systems to become more stable, flexible, cost-effective, which would help transition to cleaner and sustainable energy networks.

1.3. Research questions

This study seeks to answer the following key questions:

- What is the quantitative impact of increasing cell temperature on the electrical efficiency of photovoltaic modules?
- Which types of advanced energy storage systems are most suitable for integration with photovoltaic systems to capture and utilize waste heat?
- How can a hybrid configuration of thermal recovery and energy storage be designed to maximize overall system efficiency?
- What are the technical and economic feasibility factors associated with implementing such a system in real-world applications?

These questions aim to guide the investigation and provide a structured approach to evaluating the proposed concept.

1.4. Research hypothesis

The central hypothesis of this research is that:

"Integrating advanced energy storage systems with photovoltaic modules to capture and utilize waste heat will significantly improve the overall efficiency and economic performance of solar energy systems."

It is hypothesized that by recovering and repurposing the thermal energy that would otherwise be lost, it is possible to:

- Reduce the operating temperature of PV modules.
- Increase the net energy yield per unit area.
- Improve the financial return and sustainability of solar installations.

transmit the heat into the environment or reuse it in simple thermal processes without employing it in the advanced energy storage devices [5]. This is a lost chance of becoming overall more energy-efficient and even better to invest in solar installations cost-effectiveness.

Thus, the point to investigate new methods of converting this waste heat and finding new uses with it appears self-evident, especially in the context of linking to contemporary energy storage solutions as to supercapacitors and high-efficiency batteries.

1.2. Research significance

The importance of the research is in solving two big problems occurring in the solar-energy systems: energy intermittency and thermal losses . Presenting the possibilities of collecting the waste heat that PV modules release, the paper also helps to make solar energy systems more efficient and sustainable. To begin with, electrical efficiency and life of PV modules is positively affected by the direct control of temperature, which increases the solar projects during come out of investments [6]. Secondly, although solar energy is intermittent, it can be overcome by converting the waste heat into storable energy that can either be thermal or electrical and used to smoothen out the load and enhance the stability of the grid [7]. Besides, with the incorporation of advanced energy storage devices such as supercapacitors, hybrid solar systems which can also provide electricity and usable heat on the same system under one installation becomes another avenue. This strategy will emerge with the trend of circular energy systems and smart energy grid worldwide [8], particularly in areas with high levels of solar irradiance where overheating is a typical issue. The study, finally, helps the larger aim of enabling

In this paper, we suggest a different method of utilizing waste heat of solar cells with the help of well-developed energy storage systems, which would enhance the life cycle and performances of the ones installed PV. The paper looks at ways in incorporated thermal recovery that can offset thermal losses and improve the economics of solar power systems especially in hot climates where heat intolerance would be a recurring problem.

1.1. Problem statement

The PV solar cells are proving to be the backbone of a renewable energy system because they are scalable, environmentally friendly and cost is coming down [1]. But the problem that comes along with the PV technology is that as the temperature increases so do the adverse effects of the same on the system. Only some percentage of incident energy is converted into electricity with the remaining energy being loss to heat as the light energy hits the surface layer of a PV module [2]. This thermal build up results in marked decrease in the electrical efficiency especially in hot environments where there is high ambient temperature and the cooling systems are most times inadequate. It has been established that heating above normal operating temperatures, the efficiency of standard silicon based PV modules can reduce by around 0.3-0.5 per cent per degree Celsius [3]. This does not only minimize the power output but also increases the rate of wear and tear of the materials used and decreases the durability of the panels as time goes [4].

Although attempts have been made at controlling this problem by passive cooling or introducing the concept of hybrid photovoltaic-thermal (PVT) systems, there still exists a loophole in making good use of the so called waste heat that is produced. The majority of the existing systems either

1. Introduction

The trend in the world of aligning to the renewable form of energy sources has escalated the utilization of the photovoltaic (PV) solar systems quite high because of the importance they contribute in terms of the environment and the reduced cost [1]. Nonetheless, one of the major problems of the PV technology is the adverse effect of temperature increase on the power efficiency. Because the solar cells can absorb very little of the sunlight that hits them, only a small portion of the light is changed to electricity and the rest is deflected as waste heat [2].

Research studies have indicated that there is a decrease in electrical efficiency of crystalline silicon PV modules of about 0.3 percent to 0.5 percent due to every 1 Deg C increment in temperature above the nominal operating cell temperature (NOCT) [3]. Beside limiting power generation, this thermal degradation also causes a faster material aging and shortens the life of PV modules operational lifespan [4]. It is thus very crucial to regulate this excess heat to ensure continued performance and duration of this system.

Rather than this waste heat is an inevitable byproduct, recent research indicates that this waste heat could be harnessed and used efficiently using integrated thermal management procedures and energy storage systems [5]. As an example, hybrid photovoltaicthermal (PVT) modules operate as power slides and thermal collectors, which improves the total energy output [6].

Also, elevated energy storage equipment, including lithium-ion batteries, supercapacitors, as well as phase-change materials, are more being researched in regards to their ability of storing both electrical and thermal energy effectively [7]. Such systems allow a more favorable load balancing, grid stability and energy use, particularly off-grid or intermittent solar [8].

المستخلص

تُنتَج الطاقة المتجددة في معظمها من خلال الخلايا الشمسية الكهروضوئية (PV)، إلا أن هذا النوع من الإنتاج يتأثر بشكل ملحوظ بتراكم الحرارة أثناء عمل الخلايا. فالحرارة الزائدة لا تقتصر آثارها على خفض كفاءة الإنتاج الكهربائي، بل تُسهم كذلك في تسريع تآكل الخلايا وتقليص العمر الافتراضي لوحدات الطاقة الشمسية. في هذه الدراسة، أقترح نظامًا هجينًا يجمع بين الخلايا الشمسية عالية الكفاءة ومكثفات فائقة الأداء (Supercapacitors) تُستخدم كوحدات لتخزين الطاقة، إلى جانب استثمار الحرارة المهدورة كمصدر طاقة إضافي يُسهم في رفع كفاءة النظام بشكل متكامل. ولتقدير الإمكانات الحرارية القابلة للاسترداد، وكذلك تقييم مدى جدوى استخدام المكثفات الفائقة في توفير طاقة تخزين قصيرة الأمد و قد طورنموذج نظري شامل لهذا الغرض، الذي اعتمد على النمذجة الرياضية والتحليل المقارن لاختبار أداء النظام، مع التركيز على عناصر محددة مثل انخفاض درجة الحرارة، وتحسين الخرج الكهربائي، وتعزيز القدرة التخزينية للطاقة. وأظهرت نتائج الدراسة أن استخدام مبادل حراري في الجهة الخلفية من الخلايا الشمسية يمكن أن يخفض درجة حرارة الخلايا بما يتراوح بين 10 إلى 15 درجة مئوية، الأمر الذي ينعكس بزيادة في القدرة الكهربائية بنسبة تقديرية تتراوح ما بين 4% إلى 6%. كما أظهرت الاختبارات أن المكثفات الفائقة تمتلك سعة تخزين كافية لاحتواء كميات كبيرة من الطاقة خلال فترات زمنية قصيرة، مما يجعلها وسيلة فعّالة لتعويض التقلبات التي قد تحدث بسبب تغيّر شدة الإشعاع الشمسي. ويبرز هذا التصميم الهجين بوصفه نهجًا واعدًا لتحسين الأداءين الكهربائي والحراري، خاصة في المناطق ذات المناخ الحار، ما يعزز فرص تطوير أنظمة طاقة شمسية أكثر كفاءة واستدامة ومرونة في المستقبل، ويمكن لاحقًا التحقق من فاعليته عبر تجارب تطبيقية.

الكلمات المفتاحية: الخلايا الشمسية الكهروضوئية، استعادة الحرارة المهدورة، المكثفات الفائقة، إدارة الحرارة، أنظمة الطاقة الهجينة، تخزين الطاقة.

Abstract

Renewable energy is largely produced by photovoltaic (PV) solar cells, but such a process is greatly impacted with thermal buildup as the solar cells work. Redundant heat not only lowers the level of electrical production, but also hastens deterioration and decreases the life-time of PV modules. In this study, we suggest a hybrid system which combines high-performance energy storage devices, namely, supercapacitors, and PV modules coupled with the use of waste heat as a source of useful energy and which increases efficiency of the system in conjunction. In order to estimate the potential amount of thermal energy recovery, as well as to determine the viability of using supercapacitors to provide short-term energy storage it was created a theoretical model. To test the system performance such as temperature reduction in the system, electrical output enhancement, and energy storing capacity, mathematical modeling and comparative analysis were made. The outcomes show that with the application of a rear-side heat exchanger, the temperature of cells can be lowered up to 10 -15C, which will yield a power recovery of about 4 -6% . Also, the supercapacitors were proved to have sufficient storage capacity to hold large energy in a short period of time and could have been utilized to eliminate any variation of disruption due to variable solar irradiance. The hybrid design under consideration demonstrates great prospects to improve the electrical and thermal performance, especially within hot climates regions. This will help in the designing of more efficient, future sustainable and adaptive solar energy systems and can be used in the future to validate through experiments.

Keywords: Photovoltaic systems, Waste heat recovery, Supercapacitors, Thermal management, Hybrid energy systems, Energy storage.

"Harnessing Waste Heat from Solar Cells Using Advanced Energy Storage Systems" / Original study

Chief Engineered Hayder Ibrahim Ismael
Ministry of Higher Education and Scientific Research,
University of Mustansiriyah, Baghdad / Iraq
Haideribrahim2021@gmail.com
+9647705498521

"استثمار الحرارة المهدورة من الخلايا الشمسية عبر أنظمة تخزين الطاقة المتقدمة"

ر.مهندسين حيدر ابراهيم اسماعيل
وزارة التعليم العالي و البحث العلمي، الجامعة المستنصرية – بغداد \ العراق

[37]. Y. J. Lim *et al.*, "Advances in Mobile Biometric Security," IEEE Access, vol. 9, pp. 168451–168460, 2021.

[38]. E. Russo, "Security Analysis of ECC-Based 2FA in Mobile Apps," Mobile Security Journal, vol. 15, no. 4, pp. 245–256, 2023.

[39]. C. Huang, "Two-Factor Authentication in Financial Networks," Financial Services Security, Boston, MA: McGraw-Hill, 2021, pp. 167–183.

[40]. S. Ravi and M. Arya, "Next-Gen Authentication Methods in IoT," Journal of Network and Computer Applications, vol. 176, p. 102909, 2021.

[41]. N. J. Kim and J. H. Park, "Exploring Machine Learning in 2FA," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 8, pp. 3423–3435, 2021.

[42]. M. C. Kerr *et al.*, "Implementing 2FA in Hybrid Cloud Environments," Cloud Security Journal, vol. 22, no. 5, pp. 316–327, 2022.

[43]. W. Li *et al.*, "Scalable 2FA for Mobile Banking," Journal of Banking & Finance Security, vol. 10, no. 4, pp. 55–67, 2021.

[44]. R. Singh, "Cost-Effective Biometrics in 2FA Solutions," IEEE Security & Privacy, vol. 18, no. 6, pp. 78–85, 2020.

[45]. T. Andrews, "Device Management in Certificate-Based Authentication," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 50, no. 5, pp. 3497–3507, 2020.

[46]. C. E. Torres, "Usability Issues in Multi-Factor Authentication," International Journal of Human-Computer Interaction, vol. 37, no. 4, pp. 320–329, 2021.

[47]. J. Swartz and L. Evans, "Security Standards for App-Based 2FA," Journal of Information Systems, vol. 43, no. 1, pp. 25–37, 2020.

[48]. S. Lee and J. M. Choi, "Behavioral Analysis in Adaptive Authentication," IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 2, no. 3, pp. 269–280, 2020.

[49]. A. Walker, "Passwordless Authentication in Modern Networks," Network Security Journal, vol. 34, no. 7, pp. 19–28, 2022.

[50]. Z. T. Li and P. M. Koh, "Simulating and Securing OTPs with ECC," IEEE Transactions on Cloud Computing, vol. 8, no. 3, pp. 1274–1282, 2020.

[51]. J. S. Wang *et al.*, "Multi-Factor Authentication in Distributed Systems," IEEE Transactions on Parallel and Distributed Systems, vol. 32, no. 1, pp. 59–72, 2021.

[52]. B. Fields, "Hybrid Cryptography in 2FA," IEEE Transactions on Cloud Computing, vol. 9, no. 2, pp. 673–688, 2021.

[53]. A. Coleman and R. Davis, "The Importance of ECC in 2FA," Security and Privacy in Computing Systems, Berlin, Germany: Springer, 2021, pp. 58–73.

[54]. M. Bhattacharjee, "Biometric Trends in Secure Authentication," IEEE Access, vol. 10, pp. 32244–32256, 2022.

[55]. S. K. Singh, "Performance Metrics in Multi-Factor Authentication," International Journal of Information Management, vol. 62, p. 102439, 2022.

[20]. P. Ahmad, "Advanced Two-Factor Authentication for Healthcare IoT," Health Informatics Journal, vol. 27, no. 3, pp. 2653–2674, 2021.

[21]. M. Wazid, A. K. Das, and N. Kumar, "An Advanced Survey on Security and Privacy in IoT-Based 2FA," IEEE Communications Surveys & Tutorials, vol. 23, no. 3, pp. 1345–1383, 2021.

[22]. F. Kaabar and M. R. Farooq, "Security Frameworks in Multi-Factor Authentication Systems," Journal of Information Security Applications, vol. 64, p. 102685, 2022.

[23]. S. Davis, "Application of Biometrics in Digital Certificate-Based Authentication," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 456–465, 2021.

[24]. T. Zhu *et al.*, "Exploring SIM-Swap Vulnerabilities in 2FA Protocols," Computer Networks, vol. 197, p. 108268, 2021.

[25]. H. Y. Kim, "Challenges in Implementing Biometric and Certificate-Based 2FA in Smart Cities," IEEE Access, vol. 8, pp. 171285–171299, 2020.

[26]. M. Y. Lee and R. A. Brison, "Usability and Security in Mobile Device Authentication," Journal of Information Security and Applications, vol. 51, p. 102511, 2020.

[27]. A. B. Ali, "Examining Cyber Attacks Targeting Traditional 2FA," IEEE Security & Privacy, vol. 18, no. 6, pp. 59–65, 2020.

[28]. A. Beltran and B. C. Lynn, "Efficiency in Cryptographic Protocols for Mobile Networks," IEEE Transactions on Mobile Computing, vol. 19, no. 12, pp. 2755–2769, 2020.

[29]. R. Ortega and T. Tomic, "Security of Advanced 2FA Methods in Cloud Environments," Cloud Computing and Services Science, New York, NY: Springer, 2021, pp. 98–113.

[30]. N. Rashid *et al.*, "Elliptic Curve Cryptography for Low-Power 2FA Devices," IEEE Internet of Things Journal, vol. 8, no. 3, pp. 1601–1613, 2021.

[31]. J. Warner, "Comparative Study of OTP Methods in 2FA," Information Security Journal: A Global Perspective, vol. 30, no. 2, pp. 87–98, 2021.

[32]. T. Gligor, "Device-Based Authentication Using ECC," IEEE Transactions on Information Forensics and Security, vol. 16, pp. 3936–3946, 2021.

[33]. S. Nakamura and E. Sakurai, "Behavioral Biometrics in Adaptive Authentication," ACM Transactions on Information Systems, vol. 41, no. 2, pp. 78–96, 2023.

[34]. R. Gallo, "Modern 2FA in Industrial IoT: ECC and Beyond," IEEE Industrial Electronics Magazine, vol. 14, no. 2, pp. 23–33, 2020.

[35]. G. M. Mohammad and S. Benlamoudi, "Public Key Infrastructure and Certificate-Based Authentication," IEEE Communications Surveys & Tutorials, vol. 24, no. 1, pp. 673–690, 2022.

[36]. T. Sampson, "Digital Identity Verification with Certificates," IEEE Transactions on Information Forensics and Security, vol. 17, pp. 4829–4843, 2022.

[6]. P. Rannenberg, V. Varadharajan, and C. Weber, "Evolution of Authentication Systems: Towards Multi-Factor Authentication," in Proceedings of the 13th International Conference on Security and Privacy in Communication Networks (SecureComm 2018), Singapore, 2018, pp. 619–628.

[7]. M. Alotaibi and K. Elleithy, "User Authentication Factors: From Single-Factor to Multi-Factor Authentication," in Advances in Computer and Electrical Engineering, vol. 2, no. 1, pp. 1–15, 2021.

[8]. Yubico and Ponemon Institute, "2020 State of Password and Authentication Security Behaviors Report," 2020. [Online]. Available: https://www.yubico.com/authentication-report

[9]. Deloitte, "Digital Authentication: Moving Beyond Passwords," 2023. [Online]. Available: https://www2.deloitte.com/global/en/pages/risk/articles/digital-authentication.html.

[10]. Lifewire, "What Is an Authenticator App and How Does It Work?" [Online]. Available: https://www.lifewire.com/what-is-an-authenticator-app-5180855.

[11]. I. A. Jaddoa and A. T. Kurnaz, "Developing a Two-Factor Authentication System to Identify Vulnerabilities in Public Wi-Fi," International Journal of Scientific Trends, vol. 2, no. 7, pp. 1–6, 2023.

[12]. V. Banes, C. Ravariu, B. Appasani, and A. Srinivasulu, "A Novel Two-Factor Authentication Scheme for Increased Security in Accessing the Moodle E-Learning Platform," Applied Sciences, vol. 13, no. 9675, pp. 1–16, 2023.

[13]. K. Liu, Z. Zhou, Q. Cao, G. Xu, C. Wang, Y. Gao, W. Zeng, and G. Xu, "A Robust and Effective Two-Factor Authentication (2FA) Protocol Based on ECC for Mobile Computing," Applied Sciences, vol. 13, no. 4425, pp. 1–19, 2023.

[14]. NIST, "Digital Identity Guidelines," Special Publication 800-63B, Natl. Inst. Standards Technol., 2020.

[15]. J. T. Sample and P. J. Blythe, "Usability vs. Security in Two-Factor Authentication," Information Security Journal: A Global Perspective, vol. 26, no. 2, pp. 87–98, 2017.

[16]. J. G. Vives and K. Sanchez, "Biometric Authentication: Advances in Behavioral Biometrics," IEEE Security & Privacy Magazine, vol. 18, no. 4, pp. 48–55, 2020.

[17]. G. Ziegler and C. Allen, "Evaluating Mobile App-Based Two-Factor Authentication," in Mobile Computing and Network Security, New York, NY: Springer, 2022, pp. 232–249.

[18]. C. A. Gelin, "The Role of Elliptic Curve Cryptography in Mobile Device Security," Wireless Networks, vol. 26, no. 5, pp. 1937–1952, 2020.

[19]. D. Anand and S. Datta, "2FA in IoT Systems: Security and Challenges," Sensors, vol. 21, no. 8, pp. 2671–2679, 2021.

based 2FA—offers efficient cryptographic 2fa more suitable in mobile and low powered devices; while digital certificates enhance enhanced identity proofing mechanisms. Applications of biometric authentication make use of biological characteristics in order to reduce the possibility of unauthorised access to the system. Through the estimates proposed in the study, it is possible to learn about the best strategies to apply in adopting easy-t0-use 2FA systems that can resist modern security realities. Possible future works may involve further develop procedures that use such technologies like applied machine learning and artificial intelligence for developing more intelligent and adaptive authentication systems. Such systems could vary the needs for further authentication based on the behavioural characteristics and the relative estimated threat level and would offer the administrators convenient means to combat new threats. In doing so, this paper continues to build up the framework for resilient, scalable and dynamic network security architectures to address emerging threats in cyberspace environment.

References

[1]. SANS Institute, "SANS 2021 Password Management and Two-Factor Authentication Methods Survey," SANS Institute, 2021.

[2]. R. Reese, B. Smith, and C. Johnson, "Usability Analysis of Two-Factor Authentication Methods," in Proceedings of the 2019 ACM Conference on Human Factors in Computing Systems, Glasgow, UK, 2019, pp. 1–12.

[3]. A. Alotaibi and H. Elleithy, "Multi-Factor Authentication in Cyber Physical Systems: A State of the Art Survey," IEEE Access, vol. 7, pp. 128845–128866, 2019.

[4]. K. Alghamdi, S. Alharbi, and M. Alzahrani, "Recent Trends of Authentication Methods in Extended Reality: A Survey," Applied Sciences, vol. 13, no. 3, p. 9675, 2023.

[5]. T. Dereje and D. Anand, "Trends in Two-Factor Authentication: A Survey," in Advances in Intelligent Systems and Computing, vol. 1158, Springer, 2021, pp. 145–156.

e. **Security Policy and Management**: By and large, it is impossible to speak about effective implementation of 2FA not only in terms of technological infrastructure but it has to be tied to an organisation's security policies framework. Tokens, biometric data and digital certificates should therefore be enjoined to higher security standards so as to discourage any form of easy access. Also, organisations need to be very particular about lost tokens, compromised biometric data, or revoked certificates to ensure long-term security [55].

These issues have been discussed to help an organisation come up with a decision of implementing advanced 2FA methods in their network taking into account of the realistic security measures without compromising the usability. Such planning therefore guarantees that reinforcement of security in the network is stable and can easily be scaled as well as integrated into its architecture.

9. Conclusion

Network access control cannot be complete without Two-factor authentication (2FA) but trends like using SMS or emails for OTPs are proving ineffective against new generations of cyber threats. This paper explored other types of second factor authentication as providing better solutions for advanced attacks; these integrated Elliptic curve cryptography (ECC), certify by a digital certificate, and biometric authentication; all providing better security and more robust client models than regular systems. This analysis described a model that posited high security and complemented it with usability appropriate for networked environments. Other—including ECC

to a larger audience. As mentioned earlier and more important for scaled implementations, organisations might need to roll out a set of 2FA mechanisms customised for various users [52].

c. **Compatibility**: To avoid the interruption of new 2FA solutions the compatibility with preexisting structures should be maintained. For example, digital certificates are compatible with PKI based systems while may not be compatible with older system or less standardised systems. ECC is a relatively weaker cryptographic algorithm as compared to RSA and therefore is compatible with current generation mobile devices; however, it may need some modifications for those organisations or companies which still have archaic systems embedding only RSA algorithms. Biometric systems need the compatible hardware in the device, which may be an issue when users connect to the network via the own computers or from different places. Organisations should undertake compatibility evaluation studies before implementing the frameworks so that they can avoid many of the pitfalls [43].

d. **User Training and Support**: These new and more sophisticated 2FA methods, in particular, those that the user understandably has no understanding of (like digital certificates or biometric authentication), need to be promoted and supported. As with all novel approaches, users have to know how to go about these procedures such as hardware token handling, certificate configuration or even biometric capture. User training and user support should be provided so that user does not resist the new IT security policies and procedures [54].

8. Implementation Considerations

Applying different types of the 2FA in real-world networks should meet certain criteria such as cost, scalability and compatibility. Although techniques like ECC, digital certificates, and biometrics are highly secure, their implementation differs with the organisation's network topology, user population, and cost factor [50].

a. **Cost**: Most developed 2FA systems have a very high cost when it comes to the first stage, especially if biometric devices or hardware tokens are used. While making use of biometric systems, organisation may need to procure specialised hardware such as fingerprint scanners or facial recognition scanners; this comes with high cost of installation per each network and constant maintenance costs. In the same manner, reliance on a technique like digital certificates involves costs of obtaining the certificates and /or renewing from certifying authorities. In organisations with many users, the costs of such equipment as hardware tokens or biometric devices for equipment issuance and control must be compared with anticipated enhancement of security [51].

b. **Scalability**: Another important consideration is scalability due to the size of a particular network or the variety of users in an organisation. While testing, app and SMS OTPs are easier to scale, whilst Biometric authentication, Digital certificates may need many changes at the hardware and infrastructure level. ECC-based authentication is more beneficial here because it turns out to be computationally reasonable and ideal for portable gadgets and IoT things, indeed ideal for the networks that are expected to be open

Table 3: Usability and Security Levels of Traditional vs. Advanced Two-Factor Authentication (2FA) Methods

2FA Method	Security Level	Usability Level	Advantages	Disadvantages
SMS-Based OTP	Moderate	High	Easy to use and widely supported	Vulnerable to SIM-swapping and phishing attacks; relies on cellular network availability
Email-Based OTP	Moderate	High	Convenient, no extra device needed	Susceptible to phishing and email account compromise; relies on email security
App-Based OTP (e.g., Google Authenticator)	High	Moderate	High security, not reliant on cellular network	Requires app installation and setup; potential inconvenience if device is lost
Hardware Tokens	High	Moderate	Strong physical security layer, independent of other devices	Physical token can be lost or damaged; additional cost for tokens
Biometric Authentication	Very High	Very High	Convenient, fast, and user-friendly; hard to replicate	Privacy concerns; requires compatible device hardware; potential for false rejections
Digital Certificate-Based 2FA	Very High	Moderate	Strong identity verification, revocable access	Complex to set up and manage; cost associated with certificate authority services
Elliptic Curve Cryptography (ECC)-Based 2FA	Very High	High	Efficient for mobile and low-power devices, highly secure	Complexity in implementation, specialized knowledge required

as the level of protection, the difficulty of application, and the level of user satisfaction. This way, through achieving the right mix of security and convenience, the 2FA systems that are set up and controlled by the network administrators, are more likely to be adopted, and, thus, more effective [46].

Traditional Methods vs. Advanced Methods: As a result traditional methods are easy by the users compared to modern methods of technology, in addition they do not demand more gadgets or expert inputs hence they are convenient. However, they have a middle rated security and are prone to usual cybercrimes including phishing and SIM-swapping [47]. Although these methods allow better security through such features as cryptographic techniques or unique biological characteristics, they generally entail extra setup or compatible hardware and, as a result, can complicate the operation of the programme and cost more [48].

Balancing Security and Usability: Although there are even more sophisticated approaches, like biometric authentication, where the security level is high, but the inconvenience of use can be moderate, there can exist issues with implementation costs and usability by users. For organisations where enhanced security is necessary, the methods like ECC and digital certificates should be implemented because they occupy a higher level of protection against cyber threats [49]. They, on the other hand, are more appropriate in environments with less security protocol and therefore being more quickly accessible. Table 3 presents information on several 2FA approaches as compared to their security levels and ease of use. Text messaging, email and conventional OTPs are reasonably secure because they are vulnerable to phishing and SIM-swapping. Mechanisms like, hardware tokens, biometric, digital certificates and ECC based 2FA offer much higher 2FA security through cryptographic method, physical security or biological security.

flow chart of biometric authentication system is as shown below in Figure 4 below integrating all the steps of biometric data capturing and secure access.

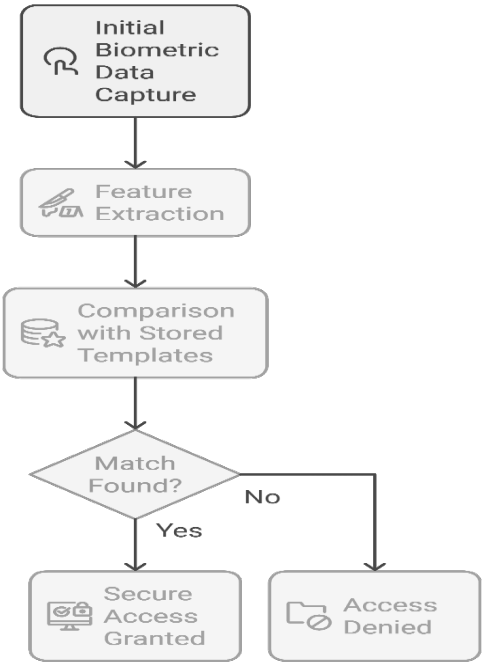


Figure 4: Flowchart of the Biometric Authentication Process.

7. Comparative Analysis of Security and Usability in 2FA Systems

The work on using 2FA in networked systems relating to this research is as follows: The degree of the security of a given networked system can be increased using 2FA while requiring the user to provide two forms of identification before gaining access to a computer. There are also enhanced forms of 2FA, like ECC, digital certificates, and biometrics, which add more layers of security, but issues of cost increase, solution complexity, and user receptiveness [44], [45]. In this section, the authors discuss a comparison of simple and current types of 2FA, analysing them based on parameters such

Biometric Authentication: Biometric authentication is a method of identification that employs the person's biological characteristics such as fingerprints, iris scans, or face. This approach provides a high level of security because a biometric measure is hard to forge. Biometric authentication is widely employed in hand-held devices, security-sensitive structures, and other security-sensitive contexts [42].

Process Steps:

a) **Biometric Enrollment**: The user initially registers their biometric data (e.g., fingerprint or facial scan) with the system, which securely stores this data as a template.

b) **User Initiates Authentication**: During subsequent logins, the user provides the biometric input.

c) **Biometric Data Capture and Comparison**: The system captures the live biometric data and compares it with the stored template to verify the user's identity.

d) **Secondary Factor Verification**: If required for 2FA, an additional factor such as a password or OTP is verified.

e) **Secure Access Granted**: Upon successful verification of both factors, access is granted.

Behavioral Authentication: Behavioural authentication is gradually being developed based on which behaviours are continuously monitored, including typing speed, the way people hold devices , and browsing history. Unlike biometric authentication which is basis on physical characteristics of an individual, behavioural authentication is instantaneous and practises the utilisation of actions from the user. This method introduces a new layer on top of 2FA, increase security at the same time keeping it user friendly [43]. A

As depicted in the following Figure 3, ECC is followed for the authentication from the generation of the authentication request up to its verification and the secure access granting. This flowchart shows how ECC transforms the conventional method through which security is implemented by using cryptographic measures in details in the detailed flowchart below.

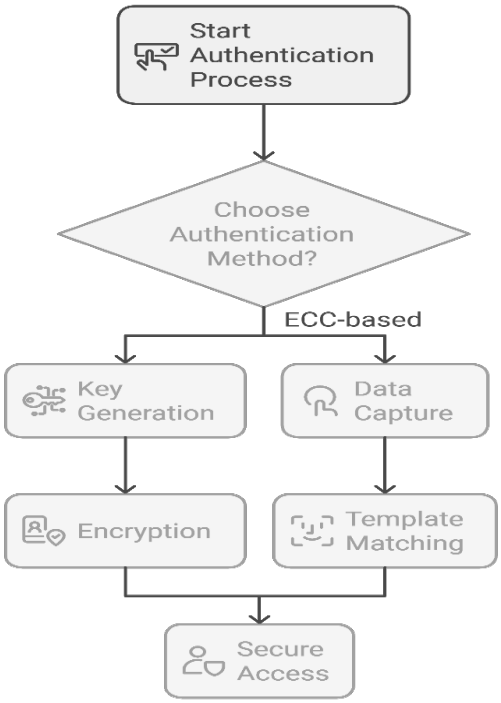


Figure 3: ECC-based authentication process.

6.2. Biometric and Behavioral Authentication in Network Security

Different biometric and behavioural techniques are powerful in the application of two-factor authentication with the use of physical characteristic features. The processes as well as benefits of biometric and behavioural authentication is also discussed, inclusive of the flowchart of biometric authentication [41].

6. ECC-Based and Biometric Authentication Process Flows

As an example of the use of the modern, ECC-based and biometrics-based, methods of 2FA, this section describes the process flows of both methods. These diagrams provide a flowchart description of each of the authentication processes, indicating the significant user-system and system-resource interactions [39].

6.1. ECC-Based Authentication Process Flow

The proposed ECC-based authentication method is effective and secure for two-factor, and ideal for environments with limited resource such as the mobile networks and the IoT application [40]. The common steps followed include key generation, data acquisition, encryption as well as template matching, for the purpose of identifying the user.

Process Steps:

a) **User Initiates Authentication**: The user begins the authentication process by providing login credentials.

b) **ECC Key Generation**: The system generates an ECC key pair for the user session, creating a digital signature based on the credentials provided.

c) **Signature Verification**: The system verifies the digital signature using the stored ECC public key to authenticate the user.

d) **Secondary Factor Verification**: An additional factor, such as a one-time password (OTP) or physical token, may be required to complete the 2FA process.

e) **Secure Access Granted**: Upon successful verification of both factors, the user is granted access to the network.

codes sent through SMS or email. Of them, the user has to submit a digital certificate with the target system, which is an encrypted papers of identification provided by the certification authority. It starts with login by the user and the system asks for a digital certificate from the user. The user then provides this certificate where the system validates for credibility or otherwise. One authenticates the certificate, and you get access; if the certificate is invalid, that means no access. This makes the method more secure as it demands both the physical holding of the certificate and confirmation by the system if needed [38]. Figure 2 A basic flowchart of use of digital certificate-based 2FA within a network system and pointing to the 'verification' step.

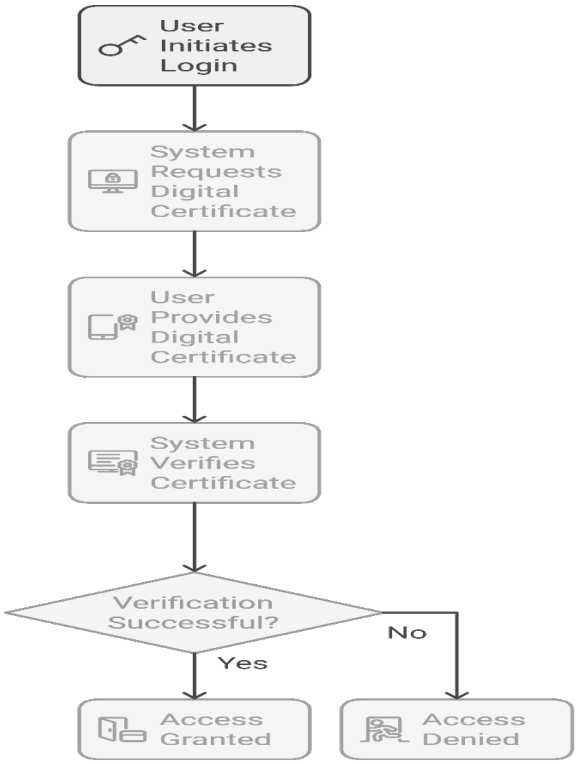


Figure 2: Digital Certificate-Based 2FA Verification Steps in a Network System.

Hence, 2FA network communication entities face threats from impersonation and privileged insider attacks. Current schemes lack advanced technical means like multifactor authentication and custom dictionaries, making them vulnerable to attacks. Researchers have attempted to design practical authentication and key agreement (AKA) protocols using hash functions and symmetric cryptography to enhance computational efficiency and communication overhead. However, these methods often lack forward secrecy for session key security. Public-key cryptography technology, such as ECC, RSA, and binary pairings, can enhance AKA protocol security but may lead to increased communication and storage costs. Balancing security and availability remains a challenge.

5. Digital Certificates and Device-Based Authentication

A digital certificate extends the security by checking the identity of user or device that is attempting to access that a network. This approach improves on more conventional 2FA by demanding users to produce a single certificate, which can be in physical form as a smartcard, or in an electronic form hosted in the cloud before being granted access [35]. The frequently used second factor is a digital certificate whereby 2FA is applied in secure environments such as e-learning platforms like moodle to secure user information [36]. With the connexion of user credentials to another standard level of certified digital identity, the previously mentioned threats and risks are minimised. Moreover, it is identified that certificate based 2FA system is flexible enough in order to be adapted in context of the limited access if the device has been compromised or of the person has changed position or job [37].

Another better performing security method is the use of digital certificate based 2FA, which are more lengthy for the second factor than the mere use of

Table 2: Comparison of Cryptographic Approaches for Two-Factor Authentication (2FA) in Terms of Security, Efficiency, and Key Size.

Cryptographic Method	Security Level	Efficiency	Key Size	Pros	Cons	Best Use Cases
Elliptic Curve Cryptography (ECC)	High (equivalent to 3072-bit RSA with a 256-bit key)	Very high; smaller key sizes require lower computational power	256-bit (for 128-bit security)	Strong security with smaller keys; efficient for mobile and IoT devices	Complex to implement, requires specialized knowledge	Ideal for mobile networks, IoT, resource-constrained environments
RSA	High (suitable for high-security applications)	Moderate; requires larger keys for similar security	3072-bit (for 128-bit security)	Widely used and supported; simpler implementation	Less efficient due to large key size, increases processing and memory demands	Suitable for secure email and VPN applications requiring high compatibility
AES (Advanced Encryption Standard)	Very high (symmetric encryption)	Very high for symmetric encryption	128-, 192-, or 256-bit keys	Fast, widely used in secure communications	Requires key exchange; not suitable for 2FA directly	Secure file transfer, cloud storage, closed systems
Digital Certificates (RSA or ECC)	High (depends on underlying algorithm)	High for user authentication; requires periodic validation	256-bit (ECC) / 2048-bit (RSA)	Strong identity verification; revocable, multi-factor compatible	Complex to manage; certification authority costs	High-security environments (e.g., banking, corporate systems)
Symmetric Key Algorithms	Moderate; faster but shared keys pose a risk	Very high, as keys are identical	Smaller (e.g., 128-bit)	Fast, efficient in closed systems	Insecure for public scenarios; key exchange required	Encrypted storage, local database security

MHE and IoT to provide data safety and privacy in achieving the performance requirements of Mobile Networks [34].

Table 2 compares various cryptographic approaches for two-factor authentication (2FA) based on three key factors: The features include security level, efficiency and size of the key or a combination of the three. ECC gives the same level of security as RSA only with matched but smaller number of bits and as such is desirable for mobile and low power applications. ECC and AES are efficient with less numbers of keys needed AES being optimised for fast Symmetric encryption. But, RSA is computationally intensive and consumes more resources especially as the key length rises. ECC provides design with a small key size enabling it to utilise the limited bandwidth and computations faster compared to a mobile and IoT devices. The benefits of each method, such as in Table 2, are also presented as well as the drawbacks, which include a high level of security, fast processing, and popularity. It thereby also provides considerations about disadvantages that may include increased complexity, challenging implementation and computational demands. ECC and AES are very appropriate for advanced uses of 2FA, while techniques such as RSA could be significantly slower due to key size and may not make efficient if used in mobile or any environment with low resource elements available.

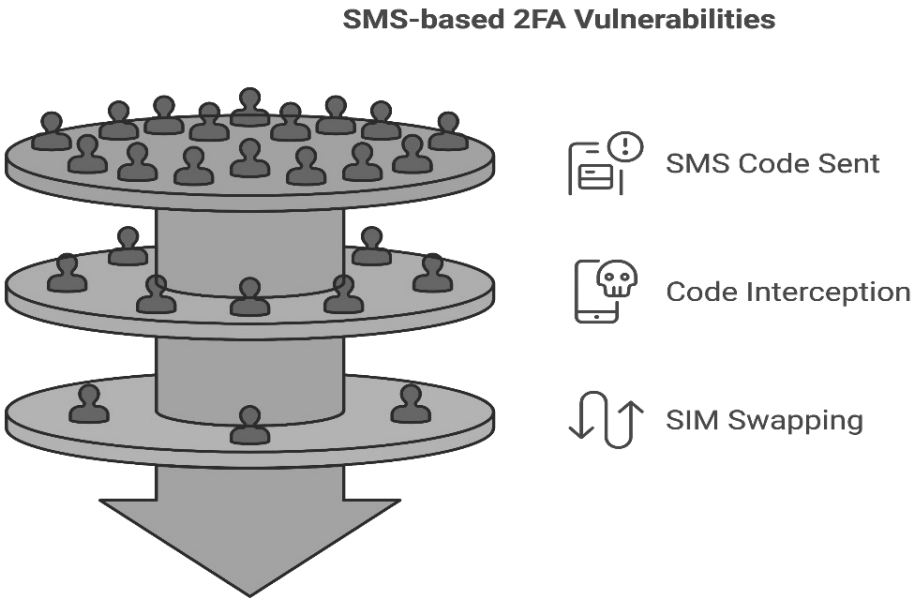


Figure 1: SMS-Based 2FA Process: Vulnerabilities to Phishing and SIM-Swapping Attacks.

4. Elliptic Curve Cryptography (ECC) for Enhanced 2FA in Mobile Networks

Elliptic curve cryptography (ECC) is a modern cryptographic technology that is getting increasingly implemented in 2FA due to its enhanced security despite it requires low resource. Mobile environment applications are well suited for ECC based 2FA because of the typical requirements in the mobile environments, for instance low resources constraints result in efficient security protocols [33]. ECC functions by producing the keys using some properties of the elliptic curves and seems to be a viable choice for a secure user authentication system. They add that ECC has a small key size, which helps to save time and memory in mobile devices while providing sufficient levels of security. For instance, ECC has been successfully incorporated in the

as they are conveyed through open communication channels with no form of secure_tunneling [28]. Moreover, they do not support smart features such as adaptive risk based authentication or device specific tokens which would be relevant to savvy, secured environment such as the financial sector or enterprise network [29]. Therefore most organisations end up in a compromise where they have to provide convenience to their users at the expense of exposing themselves to various risks affected by their systems. As a result, organisations should consider the confidentiality and ease of use when comparing traditional methods of 2FA.

Like any other popular technology, SMS-based two-factor authentication (2FA) has been compromised by a number of newer digital threats. There are two big dangers of the SMS-based 2FA; interception of the code sent and SIM-swapping. When using the SMS-based 2FA process the user receives a one-time password (OTP) via a message on their mobile device. However, this method depends on a communication channel that can easily be hacked [30]. This is so because the OTP can be intercepted through the possibly of newt work of the user or through spying on the users SMS details [31]. In addition, in SIM-swapping attack an attacker is able to mislead a mobile service provider to forward the user's phone number to an illegitimate SIM card controlled by the attacker. This makes it possible for the attacker to directly receive the OTP and therefore gains unauthorised access to the victim's account [32]. Figure 1 Simple pictorial representation of an SMS-based 2FA process and the exposed weak links to phishing and SIM-swapping attacks.

2. SIM-Swapping and Man-in-the-Middle (MitM) Attacks: SMS based 2FA is insecure due to SIM swapping and MitM where the attackers are able to penetrate established network links to compromise OTPs [24]. A SIM-swapping act involves persuading a mobile service provider to port the user's phone number to different SIM card owned by the intruder. This manipulation can be done using the social engineering and leveraging insiders at the carrier[25]. After the phone number is redirected, all SMS are sent to the attacker including OTP codes used in 2FA procedures. Worse, due to the impaired encryption in the OTP messages, the attacker can easily penetrate the 2FA layer and access the victim's accounts illicitly [26]. This kind of attack is most relevant to the current popular application of textual- and email-based 2FA, since messages are transmitted in plain text and can be easily captured without decryption. These attacks pose a major problem for services using SMS-based 2FA because the underlying communication channel is less secure than the message transmission and can easily be hijacked with an intercept or retargeting of the OTP. While modern hackers invent more effective ways to breach network vulnerabilities, conventional text message-based 2FA is not as effective [27].
3. Security vs. Usability Trade-offs: SMS- and email-based OTPs are traditional 2FA methods which have negligible barriers to adoption because of the ubiquity of phones and email [9]. Nevertheless this makes systems vulnerable to security threats, even to professional ones, at times [10]. These methods have been said to be insecure

3. Traditional Two-Factor Authentication and Limitations

The traditional approaches of two-factor authentication have been using SMS, Emails, and Time-based One-Time Password that are mobile applications. While these approaches are convenient and straightforward to implement, they face several vulnerabilities:

1. Phishing and Social Engineering: The participants described that the threat actors always advanced in utilising phishing techniques to make users vulnerable. Another type of attack is phishing where the hacker impersonates as a genuine organisation and sends emails, messages and links that looks completely genuine that makes users to enter their credentials or OTP on a fake sites controlled by the hackers [20][21]. For instance, an attacker may use email to ask the recipient to confirm their account details or to do a security cheque. The link provided is a phishing site, and once the user submits his or her credentials / OTP the attacker gets to capture and utilise this data [22]. Social engineering is much wider in its definition and in the real world it involves interacting with people directly. The attackers may come in disguises of trusted individuals, calling, emailing or even approaching the users, and force them to disclose access codes or OTPs [23]. These tactics abuse human trust and imperative, thus, it is less challenging for the attacker to 'circumvent' the second factor authentication process without sophisticated tools and methods. The classic 2FA methods that are in practise do not allow for the source of the challenge to be confirmed or for the user to confirm the identity of the sender.

Study	Methodologies Explored	Key Findings	Limitations	Proposed Improvements/ Research Gaps
Alghamdi *et al.* (2023)	Authentication in Extended Reality (XR)	Advanced methods, especially behavioral biometrics, needed for XR	Narrow application to XR environments	Investigate generalizability of behavioral biometrics beyond XR
Dereje and Anand (2021)	Trends in 2FA	Multi-layered 2FA improves security and user trust	Limited analysis of specific 2FA methods	Explore the impact of specific 2FA combinations in varying threat landscapes
Rannenberg *et al.* (2018)	Evolution of MFA	Advanced sensors enable secure, convenient authentication	No focus on emerging cryptographic methods like ECC	Investigate ECC-based authentication and other lightweight cryptography
Alotaibi and Elleithy (2021)	User Authentication Factors	Layered authentication reduces vulnerabilities	Focused on user authentication without extensive 2FA analysis	Extend findings to include comparative analysis of multi-layer 2FA methods
Yubico & Ponemon Institute (2020)	Password & Authentication Security	Highlighted risks in password practices; rising 2FA adoption to mitigate risks	Did not address advanced 2FA methods beyond traditional OTP	Explore biometric and certificate-based 2FA integration for higher security
Deloitte (2023)	Digital Authentication	Recommended MFA, biometrics, tokens over passwords	Primarily recommendations, limited empirical data	Perform empirical studies on adoption rates and practical implementation
Lifewire [19]	Authenticator Apps	Authenticator apps offer secure alternative to SMS-based 2FA	Basic overview, no in-depth analysis	Investigate scalability of app-based 2FA across various user demographics

article also elaborated that in comparison to the codes sent in SMS text messages, apps are more secure, as they do not require cell connexion and can easily be intercepted. This has been a handy practical guide, and it proffered useful experience in relation to broad adoption of the app-based 2FA in the improvement of account security [19]. The major studies associated with 2FA are summarised in Table 1 in an abbreviated manner. The individual studies are described according to its research area, findings, and methodological restrictions. It also brings out the variety of methodologies that have been used in the 2FA research including the basic techniques of SMS based verification and goes up to the complex structures of biometric and digital Certificate-based 2FA.

Table 1: Summary of Key Studies on Two-Factor Authentication (2FA) Methods

Study	Methodologies Explored	Key Findings	Limitations	Proposed Improvements/ Research Gaps
SANS Institute (2021)	Password Management, 2FA Methods	Shift towards 2FA adoption; user resistance remains a challenge	Did not analyze advanced 2FA solutions (e.g., biometrics, certificates)	Examine user education approaches for better adoption of secure 2FA methods
Reese *et al.* (2019)	SMS, Email, Phone Call, Hardware, App-based 2FA	Hardware and apps offer higher security; users prefer SMS/ email for convenience	Limited to comparing five 2FA methods	Study additional 2FA methods like biometrics and digital certificates
Alotaibi and Elleithy (2019)	MFA in Cyber-Physical Systems	Emphasized balance of security, privacy, usability	Focused primarily on cyber-physical systems	Apply findings to general networked environments; examine biometric impacts

of view. It also showed how with integration of advanced sensors and smart devices authentication can be made more secure and convenient especially in application areas that need almost constant access control [15].

In a following paper, Alotaibi and Elleithy (2021) build upon this framework by exploring user authentication factors and shift from single-factor to multi-factor in several systems. They also talked about how the use of different layers of authentication increases security and also how risks inherent with single-factor authentication such as a password or PIN are overcome, and the need for using advanced 2FA techniques [16].

The Yubico and Ponemon Institute surveyed and quantified the password and authentication security behaviours in the 2020 State of Password and Authentication Security Behaviours Report. The report then described the inherent dangers with bad password behaviour and also proposed rise in 2FA usage as an effective solution to those risks. According to the report, organisations need to look for other more reliable ways forward than passwords; for instance, biometrics authentication [17].

Digital Authentication According to the Deloitte's article released in 2023, a trend is emerging to move away from often-hackable passwords In the context of the discussed topic, passwords are considered weak links. The article suggested that multi-factor identification should be implemented, and biometric methods as well as security tokens as more secure forms and offered practical ideas on future work on authentication. This work stresses the need to develop an effective and efficient authentication process that is both highly secure and easily installable into any system.

Lastly, an article was written for Lifewire to explain the uses and the advantages of the authenticator app, which creates a TOTP for MFA. The

authentication features into aspects such as security, privacy, usability, scalability and interoperability and pointed out that a good authentication system must strive to attain a good balance of all of them. They further indicated that although multi-factor approach enhances security, its application should incorporate usability for practical implementation on real-world systems [12].

A literature review performed by Alghamdi *et al*. (2023) aimed to identify the current XR authentication methodologies and found that there was a transition from the 2FA toward methods that are context-based and include behavioural analytics along with biometrics. To this end, this work paints the picture of a challenge typical of complex technologies such as XR, where ease of authentication is always a vital component and an essential way of enhancing user experience and security, particularly when begun using the existing basic facility of the smartphone [13].

Another important piece by Dereje and Anand (2021) presented an insight into the dynamics of 2FA and focused on different ways and how efficient each of them is. Hence they stressed that it is critical to implement multiple factors in authentication in order to increase security, and thus user trust to shift from single factor to multiple factor authentication solutions. Their survey also reiterated that two-factor or many factor authentication models provide enhanced security against different cyber threats when adopted together with sturdy cryptographic mechanisms [14].

Similarly, another cross-sectional survey by Rannenberg *et al*. (2018 found) examined the development of authentication systems with special focus on the MFA. They discussed new approaches in identification and described pro and cons of biometric and behaviour-based technique as well as discussed them from both end-user and supplying-service provider points

This project aims to establish a framework for designing and implementing safe yet user-friendly two-factor authentication systems essential for countering contemporary threats while ensuring a good user experience.

2. Related Works

The analysis of two-factor authentication has been one of the main concerns in the field of cybersecurity because of its significance for securing the important data and access to accounts. Several works have focused on analysing 2FA conventional and improved techniques in terms of performance, simplicity, and issues in the current cybersecurity threat environment. Another study by the SANS Institute in 2021 focused on the current organisational practises in password management and methods used for 2FA. In this survey, an emerging trend toward its adoption for improving security with focus on the weakness of only passwords was observed. Yet the survey revealed barriers to user adoption; though 2FA enhances security, usability issues and end-user resistance are significant bi-parameters [10].

In a usability study by Reese *et al.* (2019), five common 2FA methods were evaluated: Mobile text, electronic mail, voice call, hardware token, and Smartphone application. The research confirmed that the assurance provided by the tokens of the technological hopeful was higher than with that provided by messages and emails since the latter are more fallible to phishing and SIM-swapping ploys. However, the respondents preferred the simple designs of the SMS and email-based methods, but often there are sacrifices between security and convenience [11].

In their state-of-the-art survey on multi-factor authentication in cyber-physical systems, Alotaibi and Elleithy (2019). Their research divided user

authentication have been established as viable solutions to these issues. These methods emphasize the application of recent advancements in cryptography, biometrics, and device authentication and authorization frameworks, ensuring that the utilization of secure systems remains user-friendly. For example:

1. Elliptic Curve Cryptography (ECC) is a strong cryptographic method which is particularly effective for resource-constrained devices as are the case with many IoT and mobile devices in specific [7].
2. Digital certificates bring one more level of certification in the form of verifying the identity and legitimacy of both user and accessories, which give the attacker practically no chance to mimic the identity of the user [8].
3. Fingerprint or face recognition type of authentication employs use of bodily characteristics which makes them secure and rare to be imitated [9].

These advanced 2FA techniques hold potential to provide a better solution of secure networks by expanding the base of authentication factors in knowledge-based mechanism. These methods are achieved by combining both cryptographical and biometrical parts, which make it much more difficult to enter an unauthorised code even if a password has already been become expired. In addition, such approaches have the opportunity to enhance the overall user experience by liberating users from relying on potentially dangerous or less relevant channels.

This article examines sophisticated two-factor authentication schemes, particularly the balance between security and ease, and how modern networking architecture can support the stringent security requirements of the 21st century.

1. Introduction

As the number of people online increases, the protection of one's online identity is becoming more crucial. Static passwords, which are used to secure personal information and online activities, can be easily cracked in seconds. Two-Factor Authentication (2FA) is being implemented as an extra layer of security to protect personal information and online assets. As digital and mobile connections expand, access control has become a critical component of cybersecurity. Two-factor authentication (2FA) augments authentication processes by employing two distinct components to verify a user's identity. Email and SMS-based authentications, often used as two-factor authentication (2FA), are susceptible to attacks such as phishing, SIM-jacking, and brute force methods [1]–[3].

Nonetheless, although these two-factor authentication solutions are popular, they are increasingly vulnerable to contemporary cyber attacks. Attack methods such as phishing, wherein the attacker deceives the user into disclosing information, jeopardize the integrity of email-based two-factor authentication. Similarly, SIM swapping attacks exploit weaknesses in cellular networks to gain control of the target's phone number, enabling the interception of OTPs sent over SMS [5]. These attacks expose inherent vulnerabilities in traditional two-factor authentication, as they target communication interfaces. Furthermore, even basic brute-force and social engineering attacks can compromise typical two-factor authentication systems, as users may reuse passwords or neglect standard practices [6].

The deficiencies of traditional methods of two-factor authentication underscore the necessity for adaptive, improved, and secure solutions inside networked environments. Advanced methodologies of two-factor

Abstract

With the rapid increase in cybersecurity threats targeting network systems, traditional two-factor authentication (2FA) methods are insufficient to address advanced attacks. Vulnerabilities such as phishing, SIM-swapping, and social engineering exploit the limitations of SMS-based and email-based 2FA. This paper examines advanced strategies and solutions for securing networked environments through robust 2FA mechanisms, focusing on approaches like elliptic curve cryptography (ECC), digital certificates, and biometric verification. This article offers a comparative review of various strategies about their effectiveness in enhancing security, while also highlighting their capacity to optimize user-friendliness and adaptability to emerging threats. Research findings promote an effective countermeasure strategy for network security, highlighting lessons and best practices in the implementation of advanced two-factor authentication solutions within a multi-network framework.

Keywords: -Two-Factor Authentication (2FA), Network Security, Elliptic Curve Cryptography (ECC), Biometric Authentication, Digital Certificates, Cybersecurity, Behavioral Authentication, Phishing, SIM-Swapping.

Advanced Strategies and Solutions Towards More Secure and Effective Two-Factor Authentication in Networking / Original article

Zahraa Sameer Jawad

College of Computer Science and Information Technology,
Karbala University, Iraq, Karbala
zahraa.sameer@uokerbala.edu.iq

[12] Lin, Jingzhi, *et al*. "A new steganography method for dynamic GIF images based on palette sort." Wireless Communications and Mobile Computing 2020.1 (2020): 8812087.

[13] Shu, Xinhuan, *et al*. "What makes a data-GIF understandable." IEEE Transactions on Visualization and Computer Graphics 27.2 (2020): 1492-1502.

[14] Xie, Liwenhan, *et al*. "Wakey-Wakey: Animate Text by Mimicking Characters in a GIF." Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology. 2023.

[15] Verhoeven, Geert J., Martin Wieser, and Massimiliano Carloni. "GRAPHIS—Visualise, Draw, Annotate, and Save Image Regions in Graffiti Photos." disseminate| analyse| understand gra ti-scapes (2024): 72.

[16] Hui, Shihao, *et al*. "Design and implementation of a physical layer optical fiber security communication system based on a ZUC stream cipher." Applied Optics 63.19 (2024): 5150-5158.

[17] Goulart, Ana, *et al*. "On wide-area IoT networks, lightweight security and their applications—a practical review." Electronics 11.11 (2022): 1762.

[18] Yang, Jing, Thomas Johansson, and Alexander Maximov. "Spectral analysis of ZUC-256." IACR transactions on symmetric cryptology (2020): 266-288.

[19] Tian, Hao, and Chao Wang. "A secure and lightweight implementation scheme for Internet of Things device management based on ZUC algorithm." Sixth International Conference on Computer Information Science and Application Technology (CISAT 2023). Vol. 12800. SPIE, 2023.

[20] Hoomod, Haider K., Jolan Rokan Naif, and Israa S. Ahmed. "A new intelligent hybrid encryption algorithm for IoT data based on modified PRESENT-Speck and novel 5D chaotic system." Period Eng Nat Sci 8.4 (2020): 2333-2345.

[21] Xue, Xianglian, Dongsheng Zhou, and Changjun Zhou. "New insights into the existing image encryption algorithms based on DNA coding." Plos one 15.10 (2020): e0241184.

[22] Kolivand, Hoshang, *et al*. "Image encryption techniques: A comprehensive review." Multimedia Tools and Applications (2024).

[23] Kaur, Mandeep, Surender Singh, and Manjit Kaur. "Computational image encryption techniques: a comprehensive review." Mathematical Problems in Engineering 2021.1 (2021): 5012496.

[24] Kaloshin, Vadim, and Alfonso Sorrentino. "Inverse problems and rigidity questions in billiard dynamics." Ergodic Theory and Dynamical Systems 42.3 (2022): 1023-1056.

[25] Clark, William, and Anthony Bloch. "Invariant forms in hybrid and impact systems and a taming of Zeno." Archive for Rational Mechanics and Analysis 247.2 (2023): 13.

[26] Hameedi, Balsam A., Muntaha A. Hatem, and Jamal N. Hasoon. "Dynamic Key Generation Using GWO for IoT System." JOIV: International Journal on Informatics Visualization 8.2 (2024): 819-825.

the differential attack test using NPCR and UCAI through trials. Every experiment yielded conventional findings, and the technique could be used with any kind of multimedia file format.

7. References

[1] De Azambuja, Antonio João Gonçalves, *et al*. "Artificial intelligence-based cyber security in the context of industry 4.0—a survey." Electronics 12.8 (2023): 1920.

[2] Yadav, Uma Shree, *et al*. "Security and privacy of cloud-based online social media: A survey. "Sustainable management of manufacturing systems in industry 4.0. Cham: Springer International Publishing, 2022. 213-236.

[3] Achar, Sandesh. "Cloud computing security for multi-cloud service providers: Controls and techniques in our modern threat landscape." International Journal of Computer and Systems Engineering 16.9 (2022): 379-384.

[4] Karim, Shahid, *et al*. "Current advances and future perspectives of image fusion: A comprehensive review." Information Fusion 90 (2023): 185-217.

[5] Rehman, Mujeeb Ur, *et al*. "Efficient and secure image encryption using key substitution process with discrete wavelet transform." Journal of King Saud University-Computer and Information Sciences 35.7 (2023): 101613.

[6] Evsutin, Oleg, Anna Melman, and Roman Meshcheryakov. "Digital steganography and watermarking for digital images: A review of current research directions." IEEE Access 8 (2020): 166589-166611.

[7] Jakopec, Tomislav, and Željana Hrkač. "Use of Image File Format WebP on Websites in Croatian top Domains." 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO). IEEE, 2021.

[8] Al-Sumaidaee, Ghassan, and Željko Žilić. "Sensing Data Concealment in NFTs: A Steganographic Model for Confidential Cross-Border Information Exchange." Sensors 24.4 (2024): 1264.

[9] Puchalski, Damian, *et al*. "Stegomalware detection through structural analysis of media files." Proceedings of the 15th International Conference on Availability, Reliability and Security. 2020.

[10] Mousavi, Seyyed Keyvan, *et al*. "Security of Internet of Things based on cryptographic algorithms: a survey." Wireless Networks 27.2 (2021): 1515-1555.

[11] Freire, Pedro, *et al*. "Computational complexity evaluation of neural network applications in signal processing." arXiv preprint arXiv:2206.12191 (2022).

Several objective tests are used to assess the image quality and dissimilarity between original and encrypted images, these tests include MSE, PSNR, Entropy, and SSIM. The average of these tests is explained in Table 4.

Table 4: Image quality test using a proposed method

#	Average MSE	Average PSNR	Average SNR	Average SIM	Average Entropy
1	7884.4945	0.0689	1.9019	105.6666	7.7655
2	7241.9338	0.0610	1.6410	93.6049	7.4968
3	8897.1998	0.0982	1.4129	136.1678	7.7414
4	8377.9428	0.0912	1.7823	102.7638	7.5066
5	7455.5410	0.0571	1.7707	115.4101	7.5428
6	8656.1781	0.0700	1.5366	97.5147	7.4215

8. Conclusion

Authentication, confidentiality, integrity, non-repudiation, and user privacy are some of the most crucial security requirements. These necessary safeguards for information transfer are a part of many security systems. One of the most crucial security procedures is encryption. To guarantee a high level of security, many secure encryption techniques are based on various keys and key lengths. Common files in the GIF format are used in many applications; it is crucial to secure these files during transmission. In order to generate the numbers that control the encryption outcomes, this work proposes an efficient technique for encrypting GIF files utilizing a modified version of the current algorithm, a modified ZUC stream cipher, and an efficient approach for key generation based on dynamic billiard table simulation. The suggested approach evaluates the quality of the encrypted picture, MSE, PSNR, SSIM, entropy, the time required for encryption, the unpredictability of created sequences, and

Table 3: The minimum, maximum, and mean p-value of generated seed numbers

Test #	Min. P-Value	Mean P-Value	Max. P-value	Test #	Min. P-Value	Mean P-Value	Max. P-value
1	0.0460	0.1297	0.0169	7	0.0011	0.1001	0.0012
2	0.0015	0.0315	0.0021	8	0.0063	0.1008	0.0009
3	0.1672	0.3486	0.0185	9	0.0175	0.0146	0.0004
4	0.5953	0.2725	0.1120	10	0.0021	0.0356	0.0094
5	0.1220	0.1363	0.1748	11	0.1820	0.1482	0.0385
6	0.0002	0.1016	0.0005	12	0.0042	0.1046	0.0009

From the previous table, these forms of numbers satisfy the requirements for generated keys used in the security requirements.

The complexity of any encryption algorithm depends on the technology used for encryption and decryption, and the size of the input data (frame size and number of frames) are some of the variables that can affect how long it takes to apply an encryption technique. to gauge how long it takes to encrypt and decrypt images of various sizes all frames are resized into (256x256, and 512x512) and get specific number of frames for standardization. The performance of the algorithm can then be evaluated by recording the time measurements for every image size. Table 2 measures and explains the average time to apply the suggested algorithm for the encryption and decryption of two different image sizes: (256 x 256) and (512 x 512).

Table 4: The average time-consuming for the proposed encryption/decryption algorithm

Image #	Encryption Time (256*256)	Decryption Time (256*256)	Encryption Time (512*512)	Decryption Time (512*512)
1	0.36358	0.77443	0.74989	0.11850
2	0.54773	0.38389	1.19114	0.16229
3	0.28991	1.38893	0.91124	0.11358
4	0.11385	0.91119	0.36396	0.13807
5	0.37205	0.88403	0.86212	0.05857
6	0.29301	1.39807	0.52619	0.02505

(control values), and the table dimension. Some samples of seed numbers are explained in Table 1.

Table 1: Generated numbers using a dynamic billiard table

No.	1st dimension	2nd dimension	No.	1st dimension	2nd dimension
1	0.791267678674	0.033726515230	7	0.50727098086981	0.25836350842357
2	0.523723688761	0.515946089995	8	0.42876735761504	0.48404762003481
3	0.591063640655	0.205753035517	9	0.26315682408109	0.09849427494733
4	0.266049358021	0.393284074524	10	0.32975535198079	0.12311848224074
5	0.504227801739	0.118147567635	11	0.44394351492245	0.10954989285146
6	0.251958023790	0.422023826478	12	0.29357599580899	0.10829997902249

These values from the previous table are converted to decimal values by getting the values of the digits after the floating point in fixed numbers for all values (10, 11, 12, 15, or more digits) as explained, then converted to a hexadecimal number as explained in Table 2.

Table 2: The Hexadecimal Form of Generated Numbers

No.	1st dimension	2nd dimension	No.	1st dimension	2nd dimension
1	'B83B3A99D2'	'761BB4E505'	7	'07DA41C81D'	'3C27AA46C7'
2	'79F05D4B38'	'63D486AAAF'	8	'7820C88E0B'	'70B37C87C2'
3	'899E236A4F'	'3D455E7811'	9	'2FE7D62AFD'	'16EEB75D83'
4	'3DF1C708C5'	'4CC6F35BAC'	10	'5B918EBC1C'	'1CAA6EF340'
5	'756651AA8A'	'675D18732A'	11	'1B8224D013'	'1981AEB4F3'
6	'3AA9DE566E'	'445A7E31A0'	12	'6242948C2E'	'19372E890E'

The standard NIST test is applied on the generated number after converted to binary numbers to find the randomness. Table 3. Explain the p-value of applying all tests on the generated seed numbers.

7. Experimental Results

The improved efficiency of the proposed encryption algorithm is achieved through several tests applied on a set of GIF images that are used for applications such as “robot.gif”, “cars.gif”, “airplane.gif”, “forest.gif”, “sport.gif”, and “dog.gif”. These files have different frame sizes and numbers of frames. All of them are encrypted via the proposed algorithm, and the results as explained in the following sections.

The seed number generation in the proposed method uses a dynamic billiard table to find a real number that represents the position of the ball through time. The coordination of these numbers is a two-dimensional sequence and is plotted in Figure 7 as an individual curve.

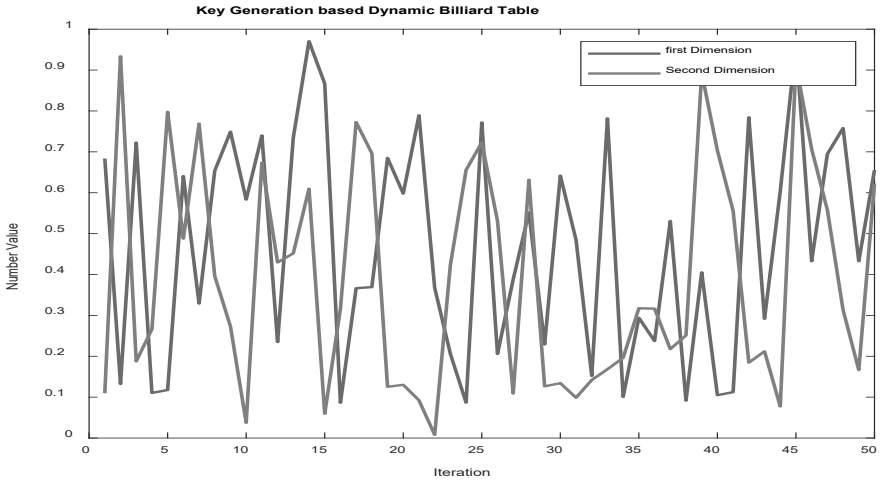


Figure 6: The Visualization of a dynamic billiard table.

The previous numbers are processed to find a hexadecimal number that is entered into the two proposed methods, Modified ZUC and Present algorithm, which need starting position (initial values) and the velocity

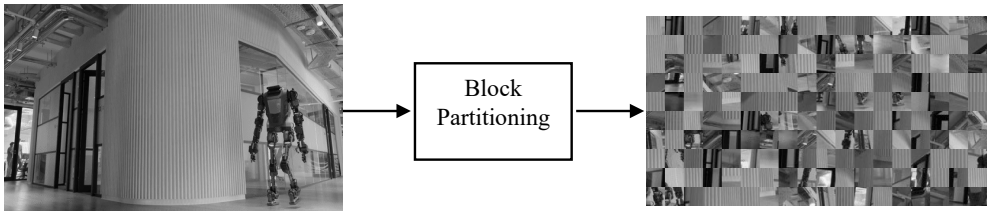


Figure 5: Partitioning of the frame into blocks process

- **Modified Present Algorithm**

The proposed method consists of two approaches for encryption; the selection of which one to apply depends on the texture complexities. If a uniform texture, a ZUC stream cipher is applied; otherwise, a modified Present algorithm is applied. The texture complexity is found by finding the variance of the block after converting it to a gray level.

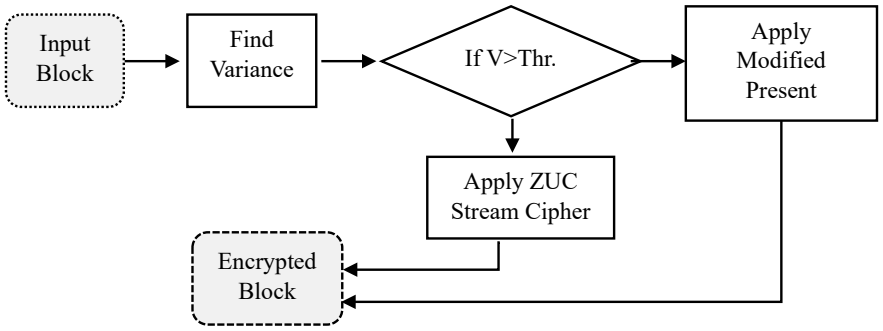


Figure 6: The proposed method for block encryption

- **Encrypted GIF reconstruction**

The encrypted blocks are merged into the matrix to reconstruct frames. The reconstruct frame's function takes the encrypted blocks and reshapes them into images using the specified frame shape. Then, the frames are merged to reconstruct the GIF file format.

Step 1: Choose the shape of the billiard table (e.g., circular, polygonal).

Step 2: Define the boundaries of the table mathematically.

Step 3: Initialize the Particle:

Step 4: Set the initial position and velocity of the particle.

Step 5: Ensure the initial position is within the boundaries of the table.

Step 6: Simulation Loop:

Update the position of the particle over time based on its velocity.

Implement collision detection to check for interactions with the walls of the table

or obstacles.

Reflect the velocity of the particle upon collision according to the laws of physics.

(elastic collisions).

Store the trajectory data (positions and velocities) at each time step.

Step 7: Generate Randomness:

Use the stored trajectory data to create a source of randomness.

This can involve sampling the position, velocity, or angles of reflection.

Convert this randomness into a suitable format for key generation.

- **Blocks Permutation**

The permutation process is done by taking the numbers generated from the proposed method for seed number generation via a dynamic billiards table. The set of selected numbers is equal to the number of blocks in the GIF frames. These numbers are sorted in ascending or descending order, and the locations of the arranged numbers are taken as an index of the location of the block that will enter the encryption process.

frame should be shown are all part of the framing process in GIF files. Simple animations that are frequently used on the web can be produced using this method. Knowing these elements will help you produce and work with GIFs programmatically. GIF file framing is explained in Figure 4.

Figure 4: GIF framing process

- **Block Partitioning of Frames**

At this stage, the image is cut into equal parts called blocks. These blocks are square and of equal size. Each of these blocks is encrypted as an independent case. The block partitioning is explained in Figure 5.

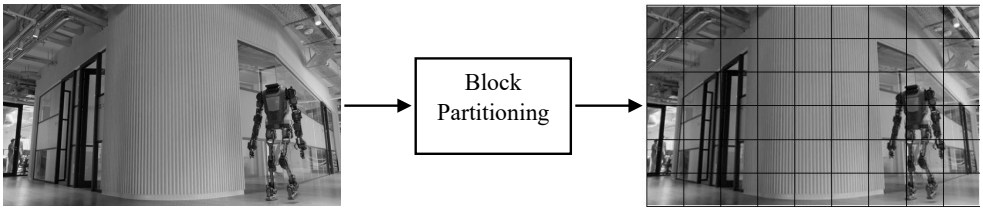


Figure 5: Partitioning of the frame into blocks process

- **Key Generation Based Dynamical Billiards**

The dynamical Billiards is used for key generation in the proposed encryption algorithm. The number generation is denoted by the following steps:

6. Proposed Method

The proposed method for encrypting a GIF image is applied by separating the frames from the GIF file. These frames are, in turn, divided into equal parts called blocks. Each block enters the encryption process. The blocks are collected in a three-dimensional matrix and permuted according to a specific key. The key generation based on dynamical billiard algorithm, the seeds of which will be used to generate keys and improve the ZUC stream cipher and Present algorithm to raise the level of its performance, then these blocks are merged into frames, and these frames are merged into a GIF file to display it as an encrypted GIF file as explain in figure 3.

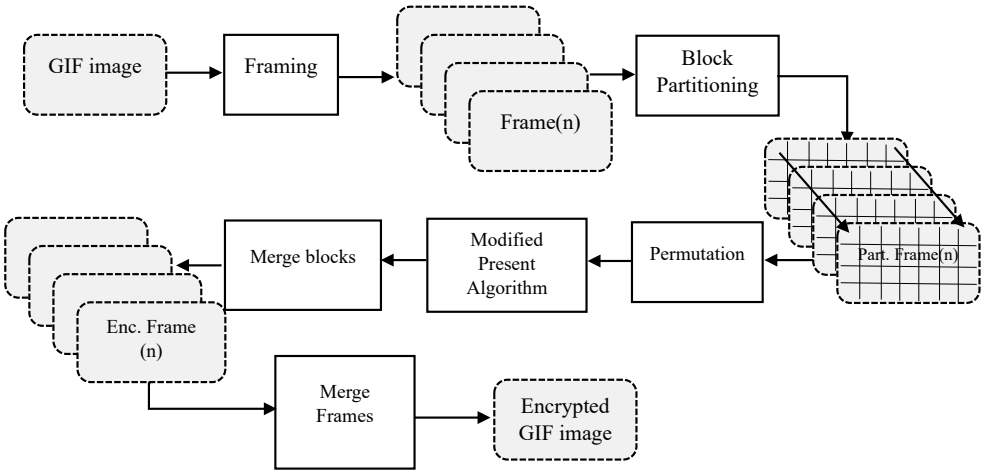


Figure 3: Present encryption algorithm

- **GIF Framing**

The framing process of the GIF file format is represented by splitting it into a sequence of slices, each one considered as an individual image. Creating many frames, controlling color tables, and specifying how each

- Including a circular key: In this phase, the round key and the state data block are subjected to a straightforward XOR operation [22].
- Substitution box layer (S-Box layer): This layer uses a corresponding Substitution Box (S-Box) to transfer a four-bit input to a four-bit output, introducing nonlinearity [23].

6. Dynamical Billiards for Key Generation

A dynamical system that represents the inertial motion of a point mass inside an area with a piecewise smooth boundary and elastic reflections is called a "dynamical billiard [24]. The angle of incidence from the border is equal to the angle of reflection. In many optical, acoustic, and classical mechanical difficulties, pool tables seem like natural models [25]. It is easy to reduce the Boltzmann gas of elastically colliding hard balls in a box—the most well-known statistical mechanics model to a pool. The theory of billiards was largely sparked by the Boltzmann-Sinai hypothesis, which has yet to be proven, regarding the ergodicity of the gas of hard balls in a torus that interact elastically. [26].

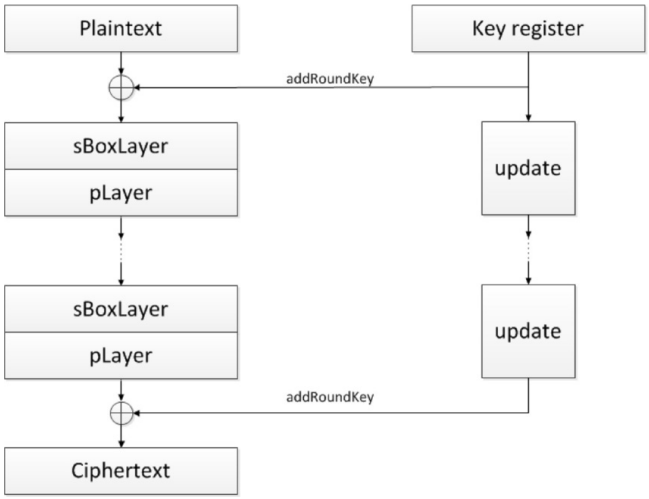


Figure 2: Present encryption algorithm

algorithms are thought to be robust and appropriate for LTE after being assessed by two more teams of distinguished specialists in addition to the algorithm standardization committee ETSI SAGE. ZUC is a stream cipher that is word-oriented [18]. A keystream Of 32-bit word is produced by using an initial vector and an initial key of 128 bits as input. It is possible to encrypt and decrypt data using this keystream [19].

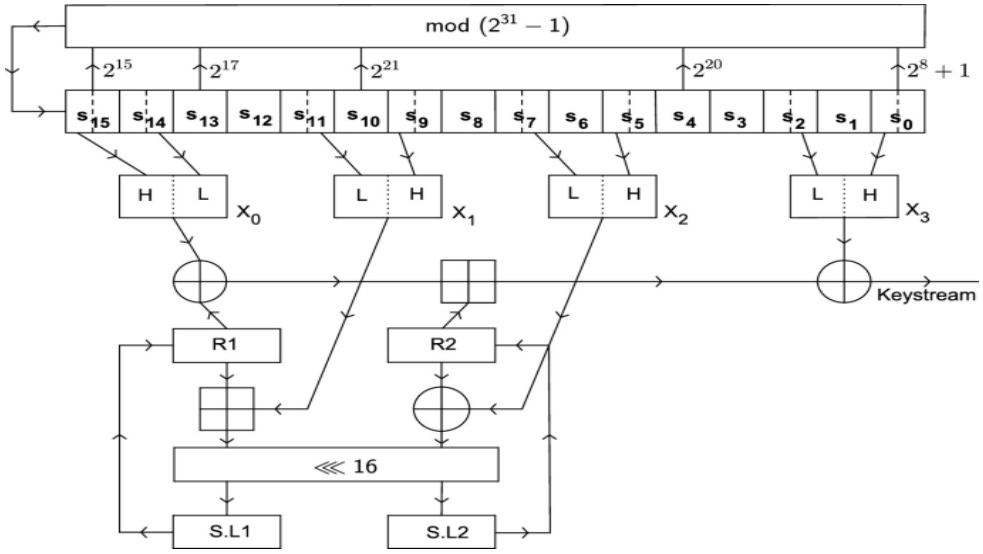


Figure 1: ZUC Stream Cipher

5. Present Encryption Algorithm

Andrey Bogdanov and associates developed the Present encryption method [20], which is a thin-film symmetric-key block cipher. It can use an 80-bit or a 128-bit key and runs on 64-bit blocks. As seen in Figure 2, Present uses a substitution permutation network topology with 31 rounds total, including a final crucial addition round [21]. Every round has these three primary purposes:

to restore the GIF image file using the present RC4 algorithm. This method is highly feasible, fast, and effective [12].

GIF Images contain a lot of redundancy, which may be used by an attacker to recover the plaintext from the ciphertext. The objective of securing Graphic Interchange Format (GIF) files by using the lightweight algorithm Present-ZUC is to design a lightweight algorithm that ensures confidentiality, integrity, and authenticity for GIF image files. Proposing a solution for an optimized and secure algorithm that can save RAM utilization by enhancing the algorithm.

2. Overview of GIF Files

A GIF (Graphics Interchange Format) file can contain multiple images or frames that are displayed in sequence to create an animation [13]. Each frame in a GIF represents a distinct image within the animation sequence [14]. Working with GIF frames involves extracting, manipulating, and possibly saving individual frames. Extracting Frames from a GIF To extract frames from a GIF, you need to read the GIF file and retrieve each frame's image data. This can be done using libraries that support GIF format manipulation [15].

4. ZUC Stream Cipher

ZUC is a secure and effective stream cipher that satisfies the requirements of contemporary communication networks. Because of its performance-focused architecture and strong security features, it is a good option for applications in wireless and mobile networks [16]. Several encryption systems, notably LTE (Long-Term Evolution) for mobile communications, use the lightweight ZUC stream cipher [17]. The ZUC

1. Introduction

The recent advances in information technology and the rise of smart devices have led to an exponential increase in the need for information exchange [1]. Images, being significant sources of information, are frequently shared over the internet by users via web browsing, social networking websites, email, messages, cloud systems, etc. [2]. However, there is a need to secure this information to avoid unauthorized access [3]. The application of conventional security techniques on images has been found inadequate due to their intrinsic properties, such as high redundancy, high correlation, high dimensionality, etc. [4]. Plain images make it easy for attackers to perform statistical analyses to retrieve crucial information [6]. Recently, digital images have been chosen as cover objects to hide secret information in steganography since the redundancy in images is very high [7]. Among various image formats, the GIF format is very popular due to its various features, such as small file size, support for animation, support for lossless compression, and usability in HTML. A lot of important and sensitive information is present in GIF images that should be protected from unauthorized access [8]. Nevertheless, GIF, being a very popular file format, is very vulnerable to attacks [9]. Many conventional algorithms have been proposed to secure the GIF images. However, conventional block ciphers and algorithms such as are not suitable for securing GIF files as they faced several limitations in computation time, key size, and security level, such as the upper limit of the key space available to possible key search, security do not guarantee chaos, not good sensitivity and so on [10]. Also, these algorithms needed complicated calculations, which increases the complexity [11]. To overcome these problems, a simple and robust technique has been proposed

Abstract

Some important security needs include authentication, confidentiality, integrity, non-repudiation, and user privacy. Many security systems include these required protections for information transmission. The encryption process is one of the most important security measures. Many secure encryption algorithms are based on different keys and key lengths to ensure a high degree of security. GIF file format is a common file format that is used in several applications, and securing these files through transmission is imperative. In this paper, an efficient method for encryption GIF files is proposed based on using the modified present algorithm, modified ZUC stream cipher, and an efficient method for key generation based on the simulation of a dynamic billiard table to generate a number that controls the encryption results. The proposed method tests through experiments on tests the randomness of generated sequences, the quality of encrypted image MSE, PSNR, SSIM, entropy, the time-consuming for encryption, and the differential attack test through NPCR and UCAI. All experiments satisfy the standard results, and the method could be used in all types of multimedia file formats.

Keywords: GIF Files. ZUC Stream Cipher, Present Algorithm

Secure GIF Files Based on ZUC Stream Cipher and Present Algorithm / Original article

Suhad Fakhri Hussein
Ministry of Education / Al-Rusafa 1, Baghdad / Iraq
suhad7242@gmail.com

[14]. Riad Chedid, “Unit sizing and control of hybrid wind-solar power systems” IEEE Transaction on energy conversion, Vol 12, No 1, March 1997.

[15]. B. Panajotovic and B. Odadzic, “Design and Inteligent” Control of Hybrid Power System in Telecommunication, IEEE Conference Melecon 2010, Valleta, April 2010.

[16]. Muhammad Aziz, Phi-Hai Trinh, Chairul Hudaya, Il-Yop Chung, “Coordinated control strategy for hybrid multi-PEMFC/BESS in a shipboard power system” Electric Power Systems Research Volume 243, June 2025, 111495

9. References

[1]. European Commision Repport, IP/09/1498, Brussels,9. October 2009

[2]. Mou Mahmood, Prangon Chowdhury, and all, "Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications" Energy Conversion and Management: X 24 (2024) 100790.

[3]. ETSI Technical Repport ETSI TR 102 532, The use of alternative energy solution in telecommunication installation.

[4]. Somalee Mitra, Basab Chakraborty, Pabitra Mitra "Smart meter data analytics applications for secure, reliable and robust grid system: Survey and future directions" Energy, Volume 289, 15 February 2024, 129920

[5]. E.F.E. Ribeiro, A.J. Marques Cardoso and C. Boccaletti "Uninterruptible energy production in standalone power systems for telecommunication", International conference on renewable energies and power quality, Valencia, April 2009.

[6]. Sophea Elmmydya Damian, Ling Ai Wong, and all, "Optimal operation of diesel generator and battery energy storage system for total fuel cost minimization in hybrid power system", Journal of Power Sources Volume 628, 1 February 2025, 235859

[7]. Reeve, "DC Power system design for telecommunication", John Wiley and Sons, USA (2007).

[8]. Gumhalter, "Power supply in telecommunications", Springer-Verlag, Berlin (1995).

[9]. D. B. Nelson, M.H. Nehrir and C. Wang, "Unit sizing of standalone hybrid wind/PV/fuel cell power generator system", IEEE Power Engineering Society Meeting 2005, Vol. 3, pp 2116-2122, June 2005.

[10]. William W. Braham, Max Hakkarainen, Munkhbayar Buyan, Gankhuyag Janjindorj, Jay Turner, Sunder Erdenekhuyag. "Cooking, Heating, Insulating Products and Services (CHIPS) for Mongolian ger: Reducing energy, cost, and indoor air pollution" Energy for Sustainable Development Volume 71, December 2022, Pages 462-479.

[11]. Karam Alhroub, Bashar Hammad, and all, "Electrical and optical characterizations and modeling of bifacial photovoltaic mini-modules integrated into solar air heaters", Solar Energy Materials and Solar Cells Volume 286, 1 July 2025, 113566

[12]. E. Muljadi and J.T. Bialasiewicz, "Hybrid power System with a ontrolled Energy Storage", 29th Annual Conference of the IEEE Industrial Electronich Society, Roanoke, Virginia, November 2003.

[13]. Seeling-Hochmuth, "Optimisation of hybrid energy systems sizing and operation control", PhD thesis, Universiti of Kassel, October 1998.

is to reduce fuel consumption, lower greenhouse gas emissions, and lower operational and logistical costs. Many algorithms can be applied to the operation of a hybrid energy system, depending on economic and technical feasibility.

This hybrid system can also be further developed by adding another type of energy storage, such as hydrogen fuel cells. This research also presents some of the investment and operating costs of the hybrid energy system components to compare economic and technical feasibility. Changes in climatic conditions and investment programs affect the cost of the system and the selection of its components, depending on the need to secure electrical power for the loads.

Overall, the purpose of this work is to provide some practical applications for hybrid energy systems, stimulate and develop new ideas in this field, and promote the widespread use of this technology, both now and in the near future.

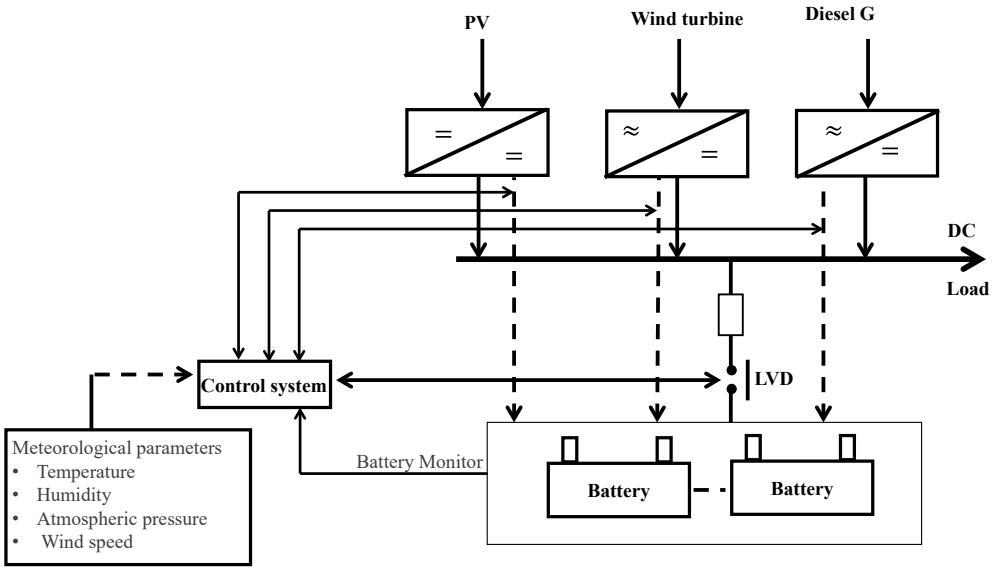


Fig5: control of the hybrid power system

8. Conclusion

These days, there is a consensus among scientists, engineers, economists, and politicians worldwide on the need to provide energy to all sectors of life, including the information and communications technology (ICT) sector, without harmful environmental impact.

In this work, a hybrid energy system is proposed that combines renewable energy sources (solar and wind) with conventional sources (diesel generators) with the possibility of storing energy through batteries. This represents a solution for the efficient use of energy resources to provide stability and reliability.

The use of renewable energy systems can be connected to the public grid (in-grid), and this system can constitute the primary source of power supply in the absence of an off-grid. Another goal of the hybrid energy system

7. Control of hybrid power system

The basic principles of hybrid power system control are based on the principles outlined in the previous paragraph. If energy is not available from renewable energy sources, it is obtained and supplied either from the battery or a diesel generator, primarily from the battery and then from the generator. The battery can be charged with a lower current, which requires a longer charging period. Conversely, if the charging current is high, the charging period is shorter, resulting in a shorter battery life [15], [16]. There are many battery parameters that must be taken into account, such as the minimum battery voltage and the maximum charging current. In this research, the variable parameters can be programmed as follows:

- Battery monitoring: Two capacitors must be calculated based on the available information: the actual capacity (representing the battery's instantaneous capacity) and the charging capacity (representing a percentage of the battery's nominal capacity). The battery can be safely used up to a charging capacity greater than the nominal capacity. If the charging capacity is close to the nominal capacity, this indicates a high charge level (increased charging time) and a low minimum allowable voltage.
- Weather information, such as air temperature, relative humidity, wind speed, and atmospheric pressure at the center's geographical location. This is to obtain accurate information over a short period.

The control system relies on the battery's state of charge, as well as controlling the diesel generator when needed. Figure (5) shows the control of the hybrid power system without the presence of a central battery charger.

6. Principle of hybrid power system operation

Figure (4) shows several types of electrical energy acquisition strategies. The primary goal of a hybrid energy system is to secure electrical energy from renewable energy sources (solar and wind energy) to power the load (communications equipment) and charge the battery, if possible (Figure (4a). If the electrical energy generated by either photovoltaic panels or a wind turbine is insufficient, the load is supplied without charging the battery (Figure (4b). When the electrical energy supply to the load from renewable energy sources (solar and wind) is interrupted, the supply is secured from the batteries and a diesel generator after the battery is discharged (Figure (4c). When the diesel generator is operating, this indicates that the battery has been discharged (4d). Care must be taken to protect the battery from deep discharge, where the minimum voltage must be fixed at a constant value $(1.8V/cell)$, and from overcharging, where the voltage must also be fixed at a constant value corresponding to the maximum current $I_{batt\max}$ ..

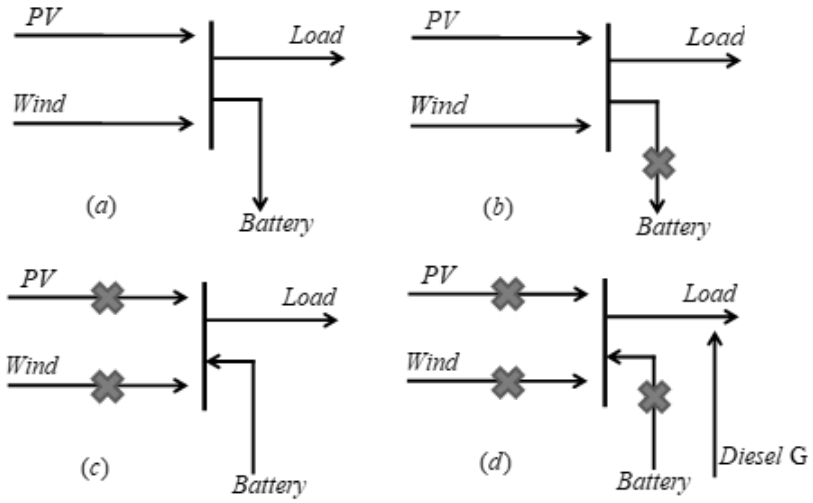


Fig 4: several types of electrical energy acquisition strategies

maximum charging current $I_{batt\max}$. Generator power is determined based on the capacity of the photovoltaic panels. From the following relationship, we find:

$$S_{\min} = \frac{P_{PV2}}{\cos\ \varphi} = \frac{9.6}{0.8} = 12kVA$$

A generator with a power of approximately $15kVA\left(\cos\phi = 0.8\right)$.

Therefore, the presence of a backup diesel generator is a positive factor for system reliability and complete power security. Typically, a large-capacity diesel generator with a sufficient power factor is recommended for powering equipment and other applications, if any. Small generators lack the high quality and capacity, and the voltage is unstable, resulting in unstable voltage and frequency. There is no possibility of using an intelligent monitoring and control system for the backup diesel generator.

Calculating the fuel tank capacity V required for the generator, assuming the generator will operate during certain times of the year (mostly in winter) $T_{aut} = 15days$. Average fuel consumption required $220g/kW$, knowing that $1Liter \approx 1Kg$.

$$V = 0.220 \times 9.6kW \times 15dayes \times 24h = 760Liters$$

Therefore, we suggest adding backup diesel generators to fuel tanks (for certain periods of the year, such as during the winter) in case electricity from the public grid is unavailable. Failure to add seasonal tanks (during the winter) increases operational and logistical costs, as well as reduces system reliability.

fuel consumption and greenhouse gas emissions. Furthermore, operational and logistical costs increase due to increased fuel tanks, services, training, and increased staffing, etc.

5-3 Power of wind turbine

The power output of a wind turbine can be determined in several ways, including the power required to supply the load only. In a hybrid power system built on solid ground, the cost of a wind turbine is lower than that of photovoltaic panels. Wind turbines are primarily dependent on wind speed. Wind turbines are used to supply telecommunications equipment and charge the battery with maximum current $I_{batt\max}$. When building a wind turbine, a wind atlas is required to obtain the optimal design, as the output voltage is affected by the wind speed in the area. The turbine is selected based on the equipment and its technical specifications.

In this paper, the turbine provides power to the communication equipment and charges the battery with maximum current $I_{batt\max} = 180A$.

Turbine capacity is determined based on the capacity of the photovoltaic panels $P_{PV2} = 9.6kW$, since the turbine's output power depends on wind speed and ambient weather conditions. Therefore, a 12kW wind turbine is selected.

5-4 Diesel back-up generators

Diesel generators are used when electrical power is not available from solar panels (due to lack of solar radiation) or a wind turbine (wind speeds are very low), and the battery is discharged. In this case, diesel generators power telecommunications equipment and charge the battery at the

5-2-2 Method II

In this method, the electrical power generated by the photovoltaic panels will be determined to supply power to the equipment and charge the battery based on the maximum current of the battery [15].

$$P_{PV2} = I_{DC}U_{DC} + I_{batt\,max}U_{DC} \quad \textbf{(13)}$$

$$P_{PV2} = 20A \times 48V + 180A \times 48V = 9.6kW$$

Therefore, the area of photovoltaic panels is determined according to this method as follows:

$$S_{PV2} = P_{PV2} / P_m^2 = \frac{9600}{135} = 71m^2$$

S_{PV2} : Photovoltaic panel area (m^2)

P_{PV2} : Power generated by photovoltaic panels according to Method II

U_{DC} : Nominal voltage 48Vdc

I_{DC} : DC current for communications equipment

$I_{batt\max}$: Maximum charging current of the battery.

Results for Photovoltaic panels Area According to the two previous methods, it was found that $S_{PV2} > S_{PV1}$: Both of the above methods were used to calculate the photovoltaic panel area. The decision to choose one of them depends on several factors, including price, weather conditions, and operational costs, among others.

In this work, we propose a larger photovoltaic panels area to reduce fuel consumption required to operate the diesel generator, lower greenhouse gas emissions CO2, and reduce logistical costs through the use of an intelligent simulation system for the hybrid energy system. In general, a smaller photovoltaic panels area can be adopted, but this leads to increased fossil

$$E_{PV} = I_{DC}U_{DC}T_{ins} + I_{DC}U_{DC}T_{back-up} \quad \textbf{(12)}$$

E_{PV}: Power generated by photovoltaic panels

I_{DC}: DC current for wired and wireless telecommunications equipment 20A

U_{DC}: Nominal voltage 48Vdc

T_{ins} : Sunshine period 9h

T_{back-p} : Battery operating time 48h

$E_{PV} = 20A \times 48V \times 9h + 20A \times 48V \times 48h = 54.720kWh$

The power required to be obtained from photovoltaic panels during the period of sunshine is expressed by the following mathematical relationship:

$$P_{PV1} = E_{PV} / T_{ins} = \frac{54.720}{9} = 6.080kW$$

To determine the optimal area for photovoltaic panels, after knowing the technical and technological characteristics of the panels (for example, a monocrystal photovoltaic panels has a higher energy density than a polycrystal cell, and the price is also higher), according to the following formula:

$$S_{PV1} = P_{PV1} / P_m^2 = \frac{6080W}{135W/m^2} = 45m^2$$

S_{PV1} : Photovoltaic panel area (m2)

P_{PV1} : Power generated by photovoltaic panels

P_m : Maximum power per unit area (m2)

5-2 Area of photovoltaic panels

Many methods and algorithms have been developed for optimal photovoltaic panels. In this study, photovoltaic panel size is a critical factor due to price and cost. The basic principle of using PV panels is to obtain electrical energy to power equipment and charge the battery. This research relies on photovoltaic panels during the summer, and the panel size is determined based on the equipment available at the center. The day will be divided into two periods: the first period 9h (during this period, the cell provides maximum power), and the second period 15h (black period, during which no electrical energy is obtained from the photovoltaic panels). The requirements of the photovoltaic panels during the first period 9h are to provide power to the telecommunications equipment and to charge the battery sufficiently during the black period 15h (the second period), which is predetermined based on meteorological data. Diesel generators can also provide power, but the goal is to reduce fuel consumption. A larger area for photovoltaic panels means more electricity, not just in the summer, but also during periods of less sunshine. The end results are linked to lower operating costs and lower greenhouse gas emissions CO_2.

Two methods are proposed to determine the size of panels necessary to provide energy to fit the available space.

5-2-1 Method I

In this method, the electrical energy generated by the photovoltaic panels during the summer will be determined to supply power to the equipment and charge the battery [9]. The energy stored in the battery is determined based on the power required to supply the equipment, according to the following relationship:

5-1 Capacity of battery

The minimum battery capacity can be calculated according to the following formula:

$$Q_{\min} = I_{DC}.T.K_1(Ah) \qquad (11)$$

I_{DC}: DC power for communications equipment (20A)

T : Battery self-recovery time

K_1 : Battery capacity factor, which depends on the type of analyzer and increases with decreasing temperature. Its value is equal to 1.15.

$$Q_{\min} = 20A \times 48h \times 1.15 = 1104(Ah)$$

In this system, four Lithium batteries (4×300Ah) connected in parallel and with the same voltage will be used 48Vdc.

$$Q = 4 \times 300(Ah) > Q(Ah)_{min} = 1104(Ah)$$

This type of battery offers a balance between price, the number of charge and discharge cycles, the depth of discharge, and the specific resistance. Another important characteristic is the high charging current, which, in turn, shortens charging times. On the other hand, a high charging current reduces battery life and increases operating costs. Typically, the battery current is fixed at a constant value, while the intelligent control system adjusts the current value based on battery condition and weather conditions.

Amperage capacity of the battery pack used:

$$C_{Ah} = 4 \times 300(Ah) = 1200(Ah)$$

Thus, the battery current according to the manufacturers:

$$I_{batt} = 0.1C_{Ah} = 120A$$

Maximum battery charging current

$$I_{batt\max} = 0.15C_{Ah} = 180A$$

- Equipment capacity (load DC) 1000W
- Load current 20A
- Need for diesel generator operation 15days/year
- Need for two batteries to operate continuously 2days

These assumptions can affect system costs, and they vary from one system to another and from one geographic region to another. In areas with greater sunshine, generator operating hours are reduced, resulting in reduced fuel consumption and environmental conservation.

5. Calculation and component sizing

The principles adopted in this work are: operation at direct current (DC) and nominal voltage 48Vdc. The inverter will not be considered in this study, given that the center's equipment operates on direct current and is powered solely by the proposed hybrid system. The remaining components operate at high efficiency. There are numerous studies and researches that take inverters (DC/AC) into account and calculate the necessary parameters (output voltage, distribution, meandering and disturbances, peak values, and harmonics). Figure (3) shows the power system used in this study and illustrated in this work.

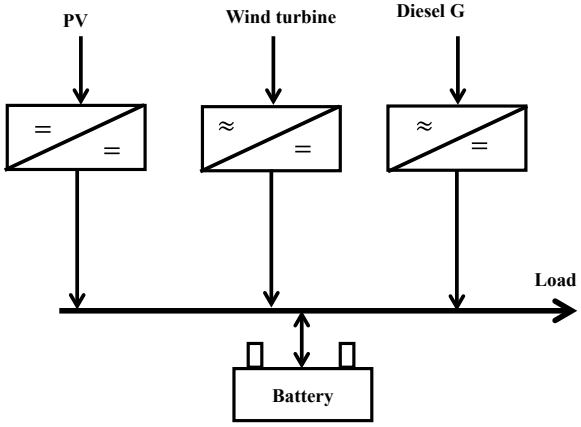


Fig 3: the power system used

where the center (operating communications equipment) or telephone booth is to be built must be known. The reliability of the system to provide electrical power must also be determined. In the detailed study of the hybrid power system, the financial costs, maintenance and replacement costs, investment and operating costs must be determined [13]. To calculate and size the power system, it is necessary to know: the geographical location, meteorological data, temperatures, the insulation period per year, the type of insulating material used, the insulation period per day, the voltage required to supply the electrical equipment, the operating time of the backup diesel generator, and the charging and discharging periods of the battery. These requirements play an important role in determining the total cost of the hybrid power system to be invested in [5].

For example, in a region with unstable and variable climatic conditions throughout the year, high-capacity batteries are used. These batteries are charged via photovoltaic panels, which are the most expensive component of the system, and also via a wind turbine. In this case, diesel generators operate at a minimum, thus reducing the fossil fuel consumption required to operate the generators. Therefore, accurate and precise data are essential when designing a hybrid energy system.

The hypotheses that will be taken into consideration in this study are:

- Temperature varies from -150C to +400C ¬
- Number of hours of sunshine per year 2400h
- Average daily sunshine 7h
- Number of hours of sunshine per day in summer 9h
- Nominal voltage 48Vdc
- Input voltages for communications equipment range from $(40.5VDC \rightarrow 57VDC)$

3-4 Diesel back-up generators

Diesel generators are used in a hybrid power system to continuously provide electrical power to telecommunications equipment. One requirement is to optimize the generator's size and reduce fuel consumption, thereby reducing carbon and greenhouse gas emissions CO_2.

It is necessary to purchase a diesel generator at the beginning of the project to calculate the costs of the diesel generator, which will later serve as a backup component in the proposed power system. We take into account the generator's capacity $R_d(kW)$ for a lifetime L_d, which is typically shorter than the turbine's or photovoltaic panel's lifetime. The initial investment value is calculated according to the following equation [14]:

$$f_d = \alpha_d . R_d . \sum_{x=1}^{x_d} \left(\frac{1+es}{1+r} \right)^{(x-1).L_d} \quad (9)$$

Which $\alpha_d(\$/kW)$ is the initial cost.

The total annual cost is as follows:

$$OM_d = \left(\alpha_{MCd} + \alpha_{OCd} \right) . R_d . T_d . fac2 \quad (10)$$

From the relationship (10), we find that the fuel price is calculated using the factor $\alpha_{OCd}(\$/kWh)$, while the operating and maintenance costs are calculated using the parameter $\alpha_{MCd}(\$/kWh)$. The total number of working hours is expressed using the factor T_d .

4. System design

Many methods and algorithms are being developed for designing and modeling hybrid power systems. To obtain the optimal model, as well as to calculate and design the system's components, the climatic data of the area

Table 1: Technical characteristics of several types of batteries.

Type	Electrolyte	Faraday effeciency %	Energy effeciency %	energy density Wh/kg	Lifespan at 20°C	Operation T$[^0C]$
Pb	H_2SO_4	>99%	75-90%	20-35%	500-2000	-20 → 60
Nickl-Cadmium	KOH	>99%	70-87%	40-60%	500-2000	-40 → 60
Nickel-metal alloys	KOH	>99%	70-87%	60-80%	500-2500	10 → 50
Lithium	$LiPF_6$	>99%	70-95%	100-200	500-4000	-20 → 60

The total cost of storing electrical energy (battery cost) is expressed by the following formula [14]:

$$f_{batt} = \alpha_b . R_b . \sum_{x=1}^{x_b} \left(\frac{1+es}{1+r} \right)^{(x-1).L_b} \quad (7)$$

α_b : Capital per kilowatt-hour

R_b : Battery cost

x_b : Number of batteries to be purchased over the lifetime of the system

L_b : Battery lifetime, which is typically less than the lifetime of the system as a whole

r : Interest or inflation rate

es : Escalation rate

Neglecting the operational cost of the battery, and taking into account the annual operating and maintenance costs $(\$/kWh/year)^{\alpha_{OMb}}$, we find that the total cost of the battery during the system lifetime is:

$$OM_b = \alpha_{OMb} . R_b . fac2 \quad (8)$$

including: the area that the turbine can occupy (A_w), the price of one square meter ($\$/m^2$) S_w ,, as well as the interest factors (r), inflation (j), and the time period of the project (system) N.

$$f_w = S_w.A_w.fac1 \quad \textbf{(4)}$$

By entering the annual maintenance cost and the annual operating cost, we can calculate the total cost over the life of the wind turbine system in terms of investment and maintenance costs.

$$OM_w = \alpha_{OMw}.A_w.fac2 \quad \textbf{(5)}$$

α_{OMw} : Annual cost of operation and maintenance

$$fac2 = \sum_{y=1}^{N} \frac{(1+es)^y}{(1+r)^y} = \frac{(1+es)}{(r-es)}.\left(1-\left(\frac{1+es}{1+r}\right)^N\right) \quad \textbf{(6)}$$

Which (*es*) is escalation rate (for *es*=r we obtain $fac2 = N$).

3-3 Energy storage

There are different types of energy storage (batteries, hydrogen fuel cells, etc.). Storage is a critical component of the energy system to ensure energy is available at all times. When selecting batteries, considerations must be given to the number of charge and discharge cycles, lifetime, technology, depth of discharge, resistance, and cost. The following table shows the technical characteristics of several types of batteries.

energy system include the small site (rooftop or adjacent garden), weather conditions, the type of cell technology, and financial costs [11].

The area occupied by the photovoltaic panels is expressed as a symbol $A_S\left(m^2\right)$, where the price per square meter is expressed as a symbol $\alpha_S\left(\$/m^2\right)$. Therefore, the cost of photovoltaic panels, after taking into account the purchase price per square meter S_S, we get:

$$f_{PV} = S_S.A_S.fac1 \quad \textbf{(2)}$$

The total annual cost after taking into account operating and maintenance costs $\alpha_{OMS}\left(\$/m^2/year\right)$ is as follows:

$$OM_{PV} = \alpha_{OMS}.A_S.fac2 \quad \textbf{(3)}$$

Where $fac1 = \dfrac{(1+j)^N}{(1+r)^N}$

r : Interest factor.

j : Inflation factor.

N : Project lifespan

3-2 Wind turbine

Electricity generation depends on wind speed, with slight changes in speed resulting in minor variations in electricity generation. On the other hand, wind disturbances result in fluctuating power generation [12]. Therefore, a key requirement is to ensure as constant and stable electricity generation as possible. When designing, a wind atlas and accurate data are required.

The operational cost of a wind turbine can be expressed by the following mathematical relationship, which is linked to several factors,

For new cabins or centers, it is preferable to build them from insulating materials according to the standard specifications used and applied in this field. In areas with high solar radiation (long periods of sunshine), it is preferable for the walls to be made of two layers of insulating materials, as shown in Figure (2). Therefore, the use of insulating materials is considered an important indicator of reduced electrical energy consumption, and on the other hand, it is considered a positive factor at the environmental level [10].

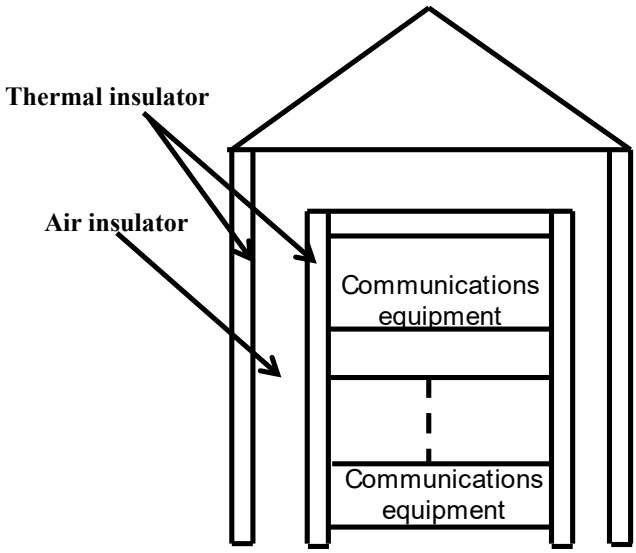


Fig 2: insulation materials

3-1- Photovoltaics panels

Solar radiation can be harnessed to generate electricity directly using a photovoltaic panel. However, the associated problem remains the high cost per watt produced by the panel. One of the key requirements in designing a hybrid energy system is optimizing the size of the photovoltaic panel. Factors that affect photovoltaic panel and must be considered when designing an

I_k : Initial value of the project (initial investment)

$S_{\mathbb{R}}$: The present value of implementing each system element

E_Y : Annual energy required by the system

OM_{pk}: Present value of operating and maintenance costs

3. Outdoor cabinet design

Some stations may be located outdoors to avoid weather factors, external influences, and mechanical protection. This requires good design for outdoor stations. Sometimes (summer, for example), the center's internal climate plays a significant role in influencing the operation of the equipment. This requires the provision of insulating materials such as rigid insulating panels or insulating glass wool (rock wool, polystyrene, polyurethane, foam, and fibricated fiber), as shown in Figure (1). The concept of thermal insulation and its applications in centers are linked to the country's legal regulations and the encouragement of investments in this field. This solution of direct insulation ensures that the internal temperature does not rise, thus reducing energy consumption.

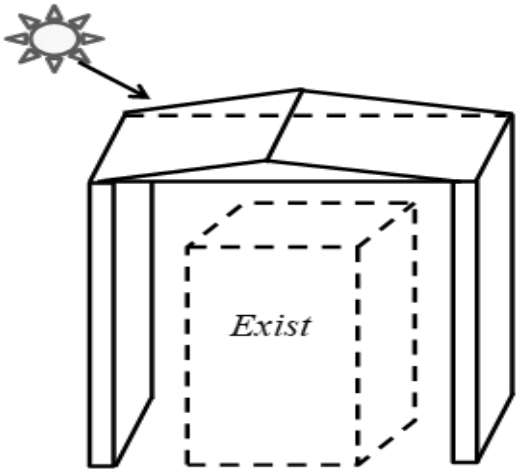


Fig 1: outdoor cabinet

a diesel generator, and a storage system (batteries). These elements play an important role in the system's operation to be at its best performance.[5], [6].

2. Requirements

The basic requirements for supplying a communications center must ensure public safety, long life, and continuous electrical power supply [7], [8].

The proposed power supply solution must achieve high efficiency in usage, energy consumption, and reliable operation of the electrical equipment connected to the center. Sometimes, the power consumed can be an important indicator of the total power consumption where the equipment is installed. The hybrid power system is designed to supply the equipment with the appropriate voltage, whether direct current, alternating current, or both. To avoid high and prohibitive costs, an optimal approach must be followed [9]. One of the design requirements is to combine costs with the potential utilization of the system, achieving both economic and technical feasibility.

When choosing renewable or hybrid energy systems, they must achieve economic feasibility within acceptable limits compared to conventional systems or obtaining energy from the public grid, in addition to the goal of preserving the environment and reducing pollution and its harmful effects.

The total cost of the hybrid energy system is expressed by the following relationship:

$$f = \left(\sum_{k=1}^{4} \left(I_k - S_{Pk} + OM_{Pk} \right) \right) \frac{1}{E_Y . N} \qquad (1)$$

k : Elements of a hybrid energy system (solar, wind, diesel generator, and battery)

N : Project lifespan (system)

1. Introduction

Many governmental and private organizations and institutions are working to utilize renewable energy and increase its efficiency in the near future, depending on available resources. Renewable energy can be used to power wired and wireless communications units using hybrid systems. Utilizing renewable energy reduces greenhouse gas emissions and lowers electricity consumption in buildings, transportation, and electrical logistics applications [1],[2]. Du to the positive benefits of renewable energy, such as mitigating global warming and other environmental issues, it also reduces the consumption of fossil fuels in electricity generation. This has prompted serious consideration of utilizing renewable energy, especially hybrid systems, to power wired and wireless communications centers in remote locations far from the public grid where electricity is unavailable [3], [4].

Renewable energy, with its high efficiency, can be a viable solution for many large investments. Therefore, governments must make regulatory, administrative, and legislative efforts and create frameworks to stimulate investments in this field due to its positive advantages. The investment proposed in this research is securing electricity for a communications center. Renewable energy sources include fuel cells, photovoltaics, wind generators, tidal, geothermal, and hydraulic. The investment proposed in this research combines renewable energy (photovoltaics and wind) with conventional energy (diesel). One of the positive features of a hybrid system (energy and storage system) is the continuous provision of electricity from both sources or one of them, thus achieving reliable electricity supply. In this research, the hybrid system will be reviewed in terms of design, calculation, size, and control. The system includes the following components: photovoltaics, a wind turbine,

Abstract

Wireless and optical communication units are becoming more widespread nowadays, especially in rural areas. Therefore, feeding electricity has become essential for the continuity of services. As a result of economic and social development, there has been an urgent need to supply electrical power for the basic requirements of wired and wireless telecommunications equipment. This is linked to public safety, long life, and connection to uninterruptible power systems to ensure continuous power supply, whether from renewable energy sources or traditional diesel systems. This research studies and designs a renewable energy (solar) power system to power telecommunications equipment. The proposed power system consists of the following main components: photovoltaic panels, a wind turbine, an energy storage system (battery), in addition to the traditional power system represented by a diesel generator. The study addresses several principles, including reliance on meteorological information and the condition of the battery used in this system. This type of system offers the following benefits: reduced fuel consumption, lower CO2 emissions, and lower logistical costs.

Design of A Hybrid System for Powering Wireless Communication Units / Original article

Samer Rabih[1, a], **Ahed Alboody**[2, b]

1 Professor at Faculty of Mechanical and Electrical Homs University Homs-Syria

2 Associate professor CESI Ecole d'ingeneurs – Nice-France

a samerrabih@gmail.com

b aalboody@cesi.fr

[43]. M. O. Hasna, "Optimization of Effective Area Spectral Efficiency for Wireless Communications Systems under Nakagami-m Fading Channels," 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall), Montreal, QC, Canada, pp. 1-6, 2016.

[44]. J. Li, Z. Song, T. Hou, J. Gao, A. Li and Z. Tang, "An RIS-Aided Interference Mitigation-Based Design for MIMO-NOMA in Cellular Networks," in IEEE Transactions on Green Communications and Networking, vol. 8, no. 1, pp. 317-329, March 2024.

[45]. E. Basar, G. C. Alexandropoulos, Y. Liu, Q. Wu, S. Jin, C. Yuen, O. A. Dobre, and R. Schober, "Reconfigurable Intelligent Surfaces for 6G: Emerging Hardware Architectures, Applications, and Open Challenges," IEEE Vehicular Technology Magazine, vol. 19, no. 3, pp. 27-47, Sept. 2024.

[46]. A. A. Yassin, M. A. Al-Husseini, and M. A. Al-Qutayri, "Reconfigurable Dual Band Antenna for 2.4 and 3.5 GHz Using Single PIN Diode," in 2013 International Conference on Computing, Electrical and Electronics Engineering (ICCEEE), pp. 63-66, 2013.

[47]. Gevez Yarkin, Tek Yusuf Islam, and Basar Ertugrul, Dynamic RIS partitioning in NOMA systems using deep reinforcement learning, Frontiers in Antennas and Propagation, Vol. 2, 2024.

[48]. X. Mu, Y. Liu, L. Guo, J. Lin and N. Al-Dhahir, "Capacity and Optimal Resource Allocation for IRS-Assisted Multi-User Communication Systems," in IEEE Transactions on Communications, vol. 69, no. 6, pp. 3771-3786, June 2021.

[49]. H. Zhou, M. Erol-Kantarci, Y. Liu and H. V. Poor, "A Survey on Model-Based, Heuristic, and Machine Learning Optimization Approaches in RIS-Aided Wireless Networks," in IEEE Communications Surveys & Tutorials, vol. 26, no. 2, pp. 781-823, Second quarter 2024.

[50]. G. C. Alexandropoulos *et al.*, "RIS-enabled smart wireless environments: Deployment scenarios network architecture bandwidth and area of influence", EURASIP J. Wireless Commun. Netw., vol. 103, pp. 1-38, Oct. 2023.

[51]. B. B. Murti, R. Hidayat and S. B. Wibowo. Spectrum Sensing Using Adaptive Threshold Based Energy Detection in Cognitive Radio System. 2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, pp. 614-617, 2021.

[52]. Vu, T.-H. *et al.* "RIS-Aided Cognitive NOMA Networks: Deep Learning Evaluation." IEEE Trans. Wireless Commun., 21(12), 10662–10677, 2022.

[53]. De Sena, A.S. *et al.* "Massive MIMO-NOMA Networks With Imperfect SIC: Design and Fairness Enhancement." IEEE Trans. Wireless Commun., 19(9), 6100–6115, 2020.

[54]. Solaiman, S. *et al.* "User Clustering and Optimized Power Allocation for D2D Communications at mmWave Underlaying MIMO-NOMA Networks." IEEE Access, 9, 57726–57742, 2021.

[55]. F. Kara and H. Kaya, "Improved User Fairness in Decode-Forward Relaying Non-Orthogonal Multiple Access Schemes With Imperfect SIC and CSI," in IEEE Access, vol. 8, pp. 97540-97556, 2020.

[32]. R. H. Yoga Perdana, T. -V. Nguyen, Y. Pramitarini, K. Shim and B. An, "Deep Learning-based Spectral Efficiency Maximization in Massive MIMO-NOMA Systems with STAR-RIS," 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Bali, Indonesia, pp. 644-649, 2023.

[33]. X. Liu, T. Wei, H. Liang, C. Guo and B. Liao, "Robust Beamforming for RIS-Assisted NOMA Systems with CSI Imperfection and Low-Resolution Phase Shifters," 2023 6th International Conference on Information Communication and Signal Processing (ICICSP), Xi'an, China, pp. 654-658, 2023.

[34]. J. Y. Baek, Y. -S. Lee and B. C. Jung, "Downlink RIS-NOMA with Constellation Adjustment for 6G Wireless Communication Systems," 2024 IEEE 21st Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, pp. 1076-1077, 2024.

[35]. J. Zhang, J. Zhang, Y. Zhou, H. Ji, J. Sun and N. Al-Dhahir, "Energy and Spectral Efficiency Tradeoff via Rate Splitting and Common Beamforming Coordination in Multicell Networks," in IEEE Transactions on Communications, vol. 68, no. 12, pp. 7719-7731, Dec. 2020.

[36]. S. Ghosh, A. Bhowmick, S. D. Roy and S. Kundu, "Outage and Ergodic Capacity Analysis in UAV-RIS Assisted NOMA-D2D Network," 2023 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), Jaipur, India, pp. 762-767, 2023.

[37]. Z. Zhu and X. Huang, "Connectivity of Wireless Networks Assisted by Transmissive Reconfigurable Intelligent Surfaces," 2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall), Hong Kong, Hong Kong, pp. 1-6, 2023.

[38]. P. Tulupov, B. Sorokin and A. Korolev, "Coverage impact of reconfigurable intelligent surfaces in 6G mobile networks," 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), Prague, Czech Republic, pp. 1-5, 2022.

[39]. M. Tavana, M. Masoudi and E. Björnson, "Dynamic RF Charging of Zero-Energy Devices via Reconfigurable Intelligent Surfaces," in IEEE Wireless Communications Letters, vol. 13, no. 8, pp. 2295-2299, Aug. 2024.

[40]. A. U. Makarfi, R. Kharel, K. M. Rabie, O. Kaiwartya, X. Li and D. -T. Do, "Reconfigurable Intelligent Surfaces based Cognitive Radio Networks," 2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Nanjing, China, pp. 1-6, 2021.

[41]. A. El Mettiti, M. Saber, A. Chehri, H. Chaibi, A. Badaoui and R. Saadane, "Reconfigurable Intelligent Surfaces and DF-relay Improved Spectral Efficiency in Cognitive Radio Networks," 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, pp. 1-6, 2023.

[42]. Y. Youn, D. An, D. Kim, M. Hwang and W. Hong, "Cognitive Reconfigurable Intelligent Surface (RIS) for mmWave Integrated Sensing and Communication," 2023 IEEE International Symposium on Antennas and Propagation (ISAP), Kuala Lumpur, Malaysia, pp. 1-2, 2023.

[19]. M. A. ElMossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han and G. Y. Li, "Reconfigurable Intelligent Surfaces for Wireless Communications: Principles, Challenges, and Opportunities," in IEEE Transactions on Cognitive Communications and Networking, vol. 6, no. 3, pp. 990-1002, Sept. 2020.

[20]. A. Araghi, M. R. Aref, M. R. A. Khandani, and M. R. A. K. M. Sadeghi, "Reconfigurable Intelligent Surface (RIS) in the Sub-6 GHz Band: Design, Implementation, and Real-World Demonstration," IEEE Access, vol. 10, pp. 2646-2655, 2022.

[21]. L. Dai, B. Wang, M. Wang, X. Yang, J. Tan, S. Bi, S. Xu, F. Yang, Z. Chen, M. Di Renzo, C. B. Chae, and L. Hanzo, "Reconfigurable Intelligent Surface-Based Wireless Communications: Antenna Design, Prototyping, and Experimental Results," IEEE Access, vol. 8, pp. 45913-45923, 2020.

[22]. S. Aboagye, A. R. Ndjiongue, T. M. N. Ngatched, O. A. Dobre and H. V. Poor, "RIS-Assisted Visible Light Communication Systems: A Tutorial," in IEEE Communications Surveys & Tutorials, vol. 25, no. 1, pp. 251-288, Firstquarter 2023.

[23]. B. Rana, S. S. Cho and I. -P. Hong, "Review Paper on Hardware of Reconfigurable Intelligent Surfaces," in IEEE Access, vol. 11, pp. 29614-29634, 2023.

[24]. Y. Mao, O. Dizdar, B. Clerckx, R. Schober, P. Popovski and H. V. Poor, "Rate-Splitting Multiple Access: Fundamentals, Survey, and Future Research Trends," in IEEE Communications Surveys & Tutorials, vol. 24, no. 4, pp. 2073-2126, Fourthquarter 2022.

[25]. H. Ge, N. Garg, A. Papaz and T. Ratnar, "A Rate-Splitting Approach for RIS-Aided Massive MIMO Networks with Transceiver Design," 2023 IEEE 24th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Shanghai, China, pp. 541-545, 2023.

[26]. G. Zheng, M. Wen, Y. Chen, Y. -C. Wu and H. V. Poor, "Rate-Splitting Multiple Access in Wireless Backhaul HetNets: A Decentralized Spectral Efficient Approach," in IEEE Transactions on Wireless Communications, vol. 23, no. 3, pp. 2413-2427, March 2024.

[27]. G. Cao, M. Li, H. Yuan, W. Chen, L. Li and A. Raouf, "Error Performance of RIS-Assisted NOMA Networks with Imperfect Channel State Information," 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, pp. 1-5, 2023.

[28]. Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network", IEEE Commun. Mag, vol. 58, no. 1, pp. 106-112, Jan. 2020.

[29]. K. Liu, Z. Zhang, L. Dai, S. Xu and F. Yang, "Active reconfigurable intelligent surface: Fully-connected or sub-connected?", IEEE Commun. Lett., vol. 26, no. 1, pp. 167-171, Jan. 2022.

[30]. T. D. Perera, D. N. K. Jayakody, S. K. Sharma, S. Chatzinotas and J. Li, "Simultaneous wireless information and power transfer (SWIPT): Recent advances and future challenges", IEEE Commun. Surveys Tuts., vol. 20, no. 1, pp. 264-302, 1st Quart. 2018.

[31]. B. Zhao, C. Zhang, W. Yi and Y. Liu, "Ergodic Rate Analysis of STAR-RIS Aided NOMA Systems," in IEEE Communications Letters, vol. 26, no. 10, pp. 2297-2301, Oct. 2022.

[8]. M. H. Babikir, K. Hamid, G. A. M. Abdu, I. Elsayed, R. A. Saeed and O. O. Khalifa, "Optimization Efficiency of 5G MIMO Cooperative Spectrum Sharing NOMA Networks," 2024 9th International Conference on Mechatronics Engineering (ICOM), Kuala Lumpur, Malaysia, pp. 183-187, 2024.

[9]. Y. Yuan, R. He, B. Ai, Z. Ma, Y. Miao, Y. Niu, J. Zhang, R. Chen, and Z. Zhong, "A 3D Geometry-Based THz Channel Model for 6G Ultra Massive MIMO Systems," IEEE Transactions on Vehicular Technology, vol. 71, no. 3, pp. 2251-2266, March 2022.

[10]. M. Hassan, R. A. Saeed, R. A. Mokhtar, K. Hamid, E. S. Ali, N. Odeh, M. Singh, O. O. Khalifa, and S. Islam, "NOMA Cooperative Spectrum Sharing Average Capacity Improvement in 5G Network," 2023 IEEE 3rd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), Benghazi, Libya, pp. 653-658, 2023.

[11]. J. Wang, C. -X. Wang, J. Huang, H. Wang and X. Gao, "A General 3D Space-Time-Frequency Non-Stationary THz Channel Model for 6G Ultra-Massive MIMO Wireless Communication Systems," in IEEE Journal on Selected Areas in Communications, vol. 39, no. 6, pp. 1576-1589, June 2021.

[12]. M. Hassan, M. Singh and K. Hamid, "Survey on NOMA and Spectrum Sharing Techniques in 5G," 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST), Nur-Sultan, Kazakhstan, pp. 1-4, 2021.

[13]. C. Wu, X. Mu, Y. Liu, X. Gu and X. Wang, "Resource Allocation in STAR-RIS-Aided Networks: OMA and NOMA," in IEEE Transactions on Wireless Communications, vol. 21, no. 9, pp. 7653-7667, Sept. 2022.

[14]. J. Chen and X. Yu, "Ergodic Rate Analysis and Phase Design of STAR-RIS Aided NOMA With Statistical CSI," in IEEE Communications Letters, vol. 26, no. 12, pp. 2889-2893, Dec. 2022.

[15]. H. Liu, G. Li, X. Li, Y. Liu, G. Huang and Z. Ding, "Effective Capacity Analysis of STAR-RIS-Assisted NOMA Networks," in IEEE Wireless Communications Letters, vol. 11, no. 9, pp. 1930-1934, Sept. 2022.

[16]. E., Basar, Di Renzo, M., De Rosny, J., Debbah, M., Alouini, M.-S., and Zhang, R. (2019). Wireless communications through reconfigurable intelligent surfaces. IEEE Access 7, 116753–116773..

[17]. Liu, Z., and Yang, L.-L. (2021). Sparse or dense: a comparative study of code-domain NOMA systems. IEEE Trans. Wirel. Commun. 20, 4768–4780, 2021.

[18]. Ohood Fadil Alwan (2023). Using Artificial Intelligence to Diagnose Demented in the Elderly, Al-Esraa University College Journal for Engineering, pp 107 – 140. V (5)-No. (8).

density mmWave environments. It adds a lot of benefits to the proposed system because it dynamically distributes power, which enables the system to adapt to user density and fluctuating network conditions, thus greatly reducing interference and resource competition. Future research will focus on significant improvements in system performance through better power allocation and user scheduling in RIS-enabled systems, in particular. The latency model is to be examined in some practical conditions, such as queuing delays, handover effects, and packet retransmissions.

References

[1]. W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The Road Towards 6G: A Comprehensive Survey," in IEEE Open Journal of the Communications Society, vol. 2, 2021, pp. 334-366.

[2]. F. Akyildiz, A. Kak, and S. Nie, "6G and Beyond: The Future of Wireless Communications Systems," in IEEE Access, vol. 8, pp.133995-134030, 2020.

[3]. F. Guo, F. R. Yu, H. Zhang, X. Li, H. Ji, and V. C. M. Leung, "Enabling Massive IoT Toward 6G: A Comprehensive Survey," in IEEE Internet of Things Journal, vol. 8, no. 15, pp. 11891-11915, 1 Aug. 2021.

[4]. M. Hassan, R. A. Saeed, R. A. Mokhtar, K. Hamid, E. S. Ali, N. Odeh, M. Singh, O. O. Khalifa, and S. Islam, "BER Improvement of Cooperative Spectrum Sharing of NOMA in 5G Network," in IEEE Access, vol. 3rd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), pp. 647-652, 2023.

[5]. H. Ahmed, A. Thair Al-Heety, B. Al-Khateeb, and A. H. Mohammed, "Energy Efficiency in 5G Massive MIMO for Mobile Wireless Network," 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, pp. 1-6, 2020.

[6]. L. You, J. Xiong, A. Zappone, W. Wang, and X. Gao, "Spectral Efficiency and Energy Efficiency Tradeoff in Massive MIMO Downlink Transmission with Statistical CSIT," in IEEE Transactions on Signal Processing, vol. 68, pp. 2645-2659, 2020.

[7]. N. Amani, S. Parsaeefard, and H. Yanikomeroglu, "Multi-Objective Energy Efficient Resource Allocation in Massive Multiple Input Multiple Output-Aided Heterogeneous Cloud Radio Access Networks," in IEEE Access, vol. 11, pp. 33480-33497, 2023.

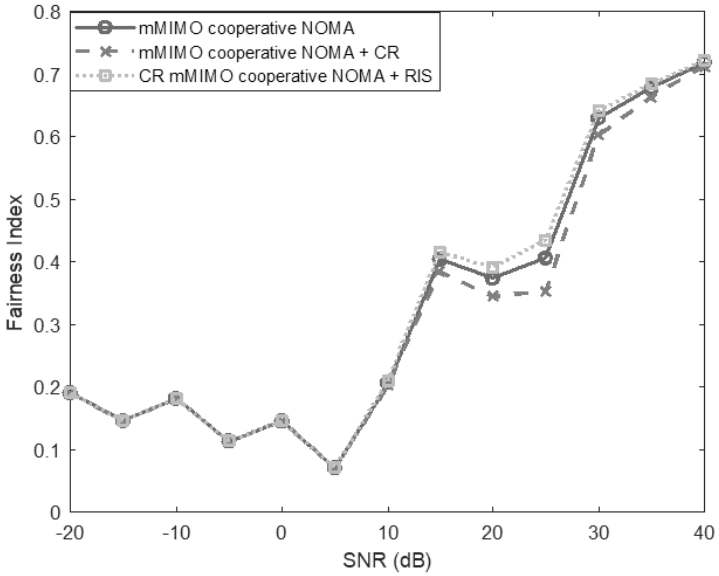


Fig. 12. Fairness index versus against SNR for three different scenarios with optimization technique

4. Conclusions

The proposed system achieves substantial enhancements in SE and throughput via the implementation of CR spectrum detection and intelligent phase shift optimization across diverse user configurations inside the mmWave environment. Studies show that adding RIS to CR-based mMIMO cooperative NOMA systems greatly lowers latency and packet loss while raising the fairness index, especially when there are a lot of users. The findings indicate that RIS-assisted cooperative NOMA is an optimal selection for future 6G networks, as it diminishes interference, enhances signal quality, and optimizes system performance under high mobility. The study's findings confirm the significant improvement and increasing importance of Q-learning approaches in reducing latency, packet loss, and the fairness index in high-

Observe the variation of the fairness index as a function of SNR across three distinct systems in Figure 12: mMIMO cooperative NOMA, mMIMO CR cooperative NOMA, and mMIMO CR cooperative NOMA with intelligent RIS, and systems utilize the Q-learning approach. When the SNR is low, particularly at 5 dB, the system becomes unstable as it fails to provide equitable power distribution among users under adverse signal conditions, resulting in the loss of advantages from CR and RIS owing to significant interference. Additionally, Q-learning may struggle to balance exploration, seeking superior solutions, and exploitation by selecting the most effective known action, resulting in instability and suboptimal fairness.

With 100 users and 40 dB SNR, the fairness index performance levels for the three systems are 0.71851, 0.71214, and 0.72123, respectively. When comparing the performance of the fairness index ratio before and after using Q-learning, we find that the fairness index improvement rate for the three systems reached 20.9%, 20.4%, and 20%, respectively. Q-learning and RIS provide an effective approach to enhance fairness in cooperative NOMA systems by dynamically distributing power and optimizing signal conditions for weaker users, hence providing a more equitable performance across all users. These outcomes surpass those of [50-55].

the NOMA cooperative mMIMO CR system, and the mMIMO cooperative NOMA CR system combined with intelligent RIS. The system is rather unstable in terms of fairness across different configurations and SNR levels, with fairness values that don't change much. CR and RIS seem to have little effect on justice, since they don't change much from the basic cooperative NOMA system. Still, they have certain benefits, including higher signal quality and a better user experience. Also, fairness is constrained by things like how power is shared, who uses the channel, and how the channel is set up, which don't change much between different setups. The performance of the three systems is approximately equal to the SNR of -15 dB and then begins to diverge. At an SNR of 40 dB, the fairness index ratings for the three systems are 0.47, 0.658, and 0.48, respectively, with 100 users.

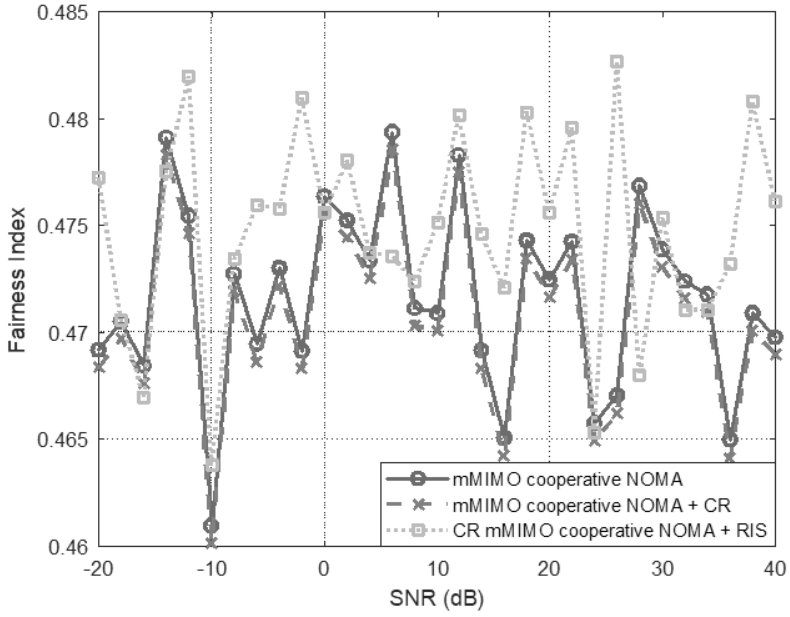


Fig. 11. Fairness index versus SNR for three different scenarios

NOMA, and mMIMO CR cooperative NOMA with intelligent RIS integration with Q-learning logarithm. At a user density of 50, the packet loss performance of three systems is 9%, 12%, and 1%, and at the user density of 100, the packet loss performance of three systems is 17%, 18.5%, and 13%, respectively, at the SNR of 40 dB. The results show that utilizing Q-learning in conjunction with RIS greatly simplifies the problem of dealing with packet loss in dense cooperative NOMA systems. This method drastically reduces packet loss rates in situations with a high concentration of users. These results surpass those of [44-54].

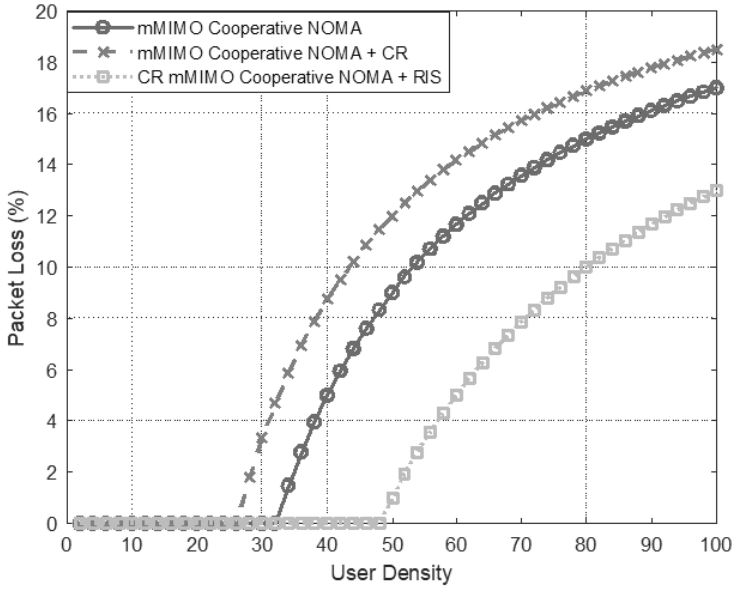


Fig. 10. Packet loss vs. user density for three different scenarios with Q-learning logarithm

Figure 11 displays the fluctuation in the fairness index in relation to SNR across three separate systems: the NOMA cooperative mMIMO system,

Figure 9 shows how packet loss varies with the number of users in three different systems: mMIMO cooperative NOMA, mMIMO CR cooperative NOMA, and mMIMO CR cooperative NOMA with intelligent RIS. At a user density of 50, the packet loss performance of three systems is 22%, 22.5%, and 21%, and at the user density of 100, the packet loss performance of three systems is 23.5%, 23.75%, and 22%, respectively, at the SNR of 40 dB. An increase in users results in greater interference and diminished resources, adversely affecting all systems. Optimal placement of RISs, phase adjustments, and additional strategies, such as user clustering and improved power distribution, can optimize performance.

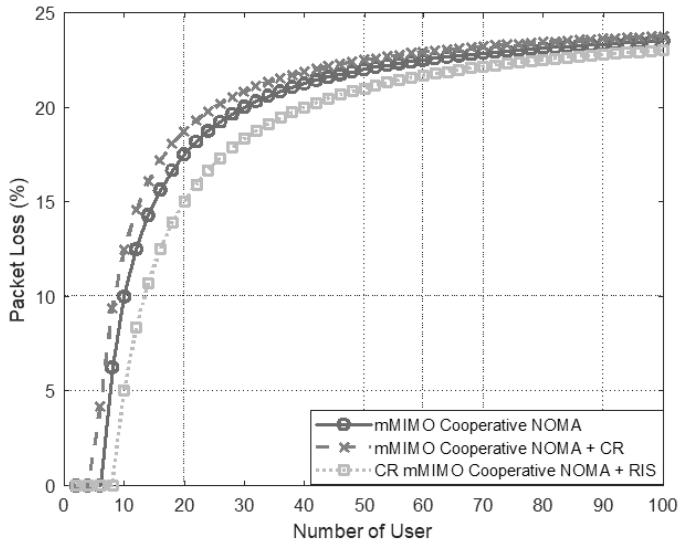


Fig. 9. Packet loss against user density for three different scenarios with optimization method

Figure 10 shows how packet loss changes with the number of users in three different systems: mMIMO cooperative NOMA, mMIMO CR cooperative

Figure 8 exhibits the variation in latency relative to the number of users across three distinct systems: mMIMO cooperative NOMA, mMIMO CR cooperative NOMA, and mMIMO CR cooperative NOMA integrated with intelligent RIS utilizing a Q-learning algorithm. At a user density of 52, the latencies recorded are 0.21, 0.23, and 0.20 ms. With 100 users, the latencies for mMIMO cooperative NOMA, CR mMIMO cooperative NOMA, and CR mMIMO cooperative NOMA systems with the intelligent RIS of 40 dB SNR are 0.47, 0.51, and 0.45 ms, respectively. These results are better than those of [44-53]. Compared to the previous result, we find an improvement of 38.2%, 37.8%, and 38.4% for three scenarios due to a reduction in latency due to the use of the Q-learning algorithm for diverse power allocation strategies based on the variation of user density, which leads to greater stability in access time.

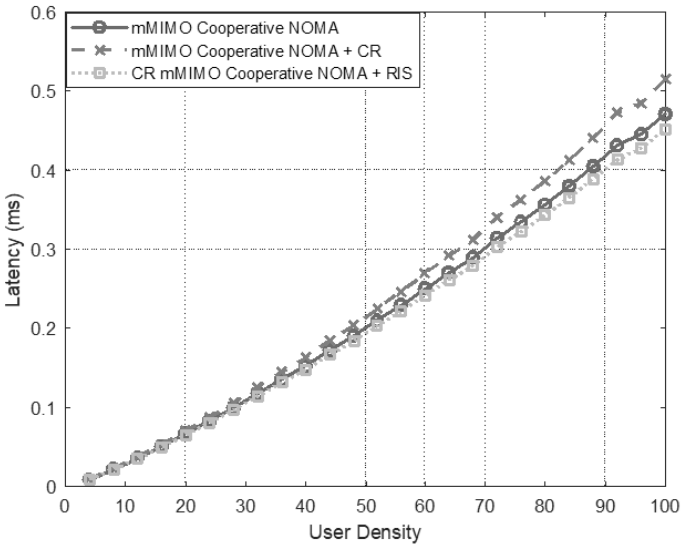


Fig. 8. Latency versus user density for three different scenarios with optimization algorithm

The latency performance of mMIMO cooperative NOMA and mMIMO CR cooperative NOMA using intelligent RIS systems was 6.8% and 5% better at 50 users, respectively, than mMIMO CR cooperative NOMA, which had a latency of 0.31 ms. At a user density of 100, mMIMO cooperative NOMA and mMIMO CR cooperative NOMA using intelligent RIS systems had 9.5% and 6% better latency performance than mMIMO CR cooperative NOMA, which had a latency of 0.87 ms at the SNR of 40 dB; the outcomes surpass those of [44]. The mMIMO CR cooperative NOMA with an intelligent RIS system has superior latency performance, even at elevated user densities, positioning it as a formidable option for high-density 6G networks where interference control is essential.

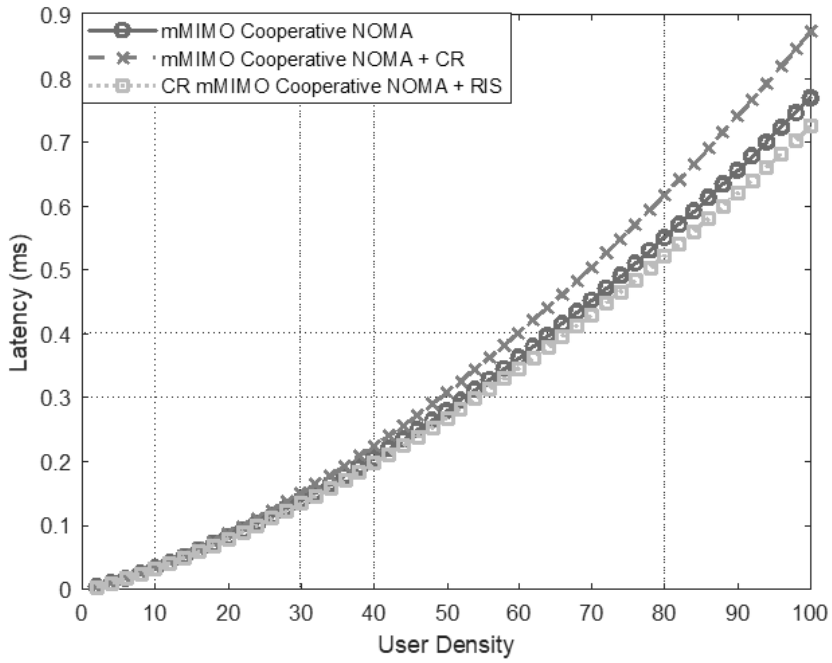


Fig. 7. Latency vs. user density for three various systems

the impact of user density by improving channel conditions and reducing interference, especially in high-user-density scenarios.

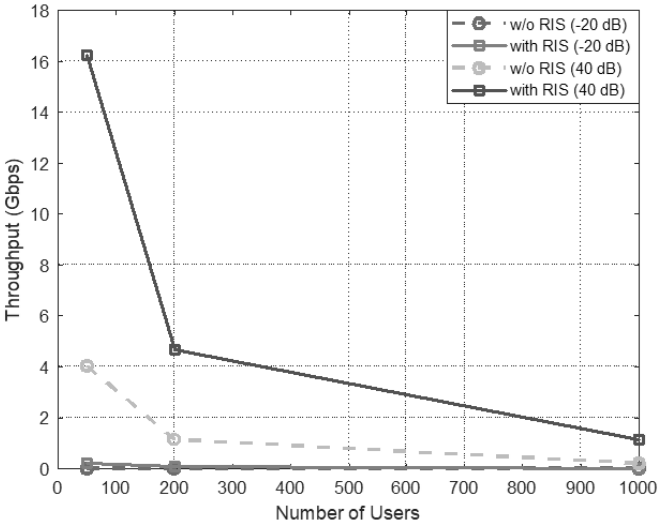


Fig. 6. Throughput versus user density of mMIMO DL CR cooperative NOMA users with and without intelligent RIS

Figure 7 illustrates how latency varies with the number of users in three systems: mMIMO cooperative NOMA, mMIMO CR cooperative NOMA, and mMIMO CR cooperative NOMA with intelligent RIS. As user competition for limited resources intensifies, network congestion and interference occur, leading to increased latency with more user density. mMIMO CR cooperative NOMA regularly exhibits higher latency than traditional NOMA utilizing the mMIMO method and mMIMO CR cooperative NOMA with RIS [52]. The heightened latency is due to the requirement to implement a coordination layer and perform channel sensing, which introduces further latencies from interference management and channel access.

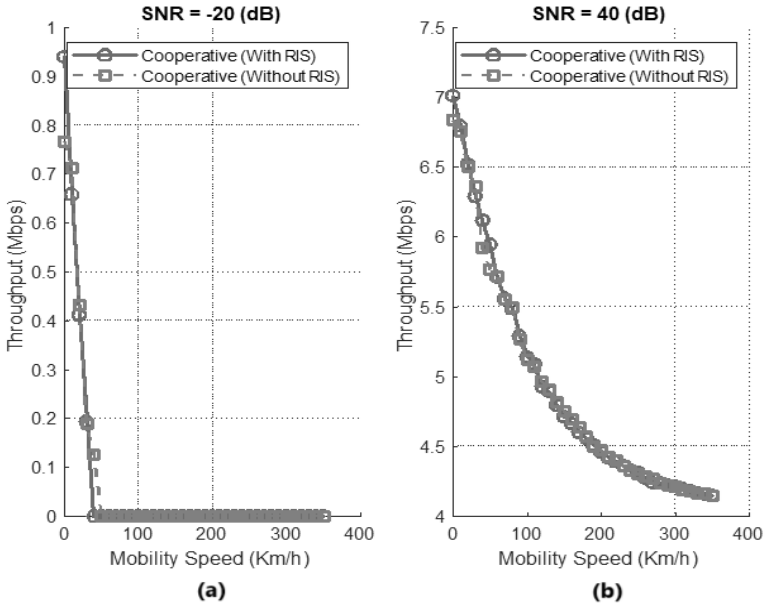


Fig. 5. Throughput against mobility speed for 100 mMIMO DL CR cooperative NOMA users with and without intelligent RIS

Figure 6 shows the mmWave mMIMO DL CR cooperative NOMA PD system with and without intelligent RIS. It shows the throughput and user density for different user numbers. Throughput reduces with the rising user density. The group of 1000 users achieved throughputs of 0.01 and 0.05 Gbps/Hz, as well as 0.21 and 1.1 Gbps/Hz, without and with intelligent RIS, at the SNR of -20 and 40 dB, respectively. A cohort of 50 users achieved throughputs of 0.1 and 0.33 Gbps/Hz, as well as 3.76 and 17.51 Gbps/Hz, without and with intelligent RIS, at SNR levels of -20 and 40 dB, respectively. The findings are better compared to those in [49].

As the user base expands, performance generally declines due to user interference and resource allocation. In cooperative NOMA systems, it is common for multiple users to compete for identical resources. RIS mitigates

Figure 5a-b shows the throughput and mobility speed (from 0 to 350 km/h) for 100 users in the cooperative NOMA PD system with mmWave mMIMO DL CR technology, with and without the intelligent RIS. As speeds increased, the throughput decreased due to Rayleigh multipath vanishing.

Figure 5-a demonstrates that a system with a (-20 dB) SNR undergoes a more rapid decrease in throughput when mobility speed escalates, unlike the system using RIS. Mobility rates of 50 km/h intensify system attenuation, resulting in a significant reduction in throughput. The poor SNR causes the signal quality to deteriorate to such an extent that augmentation of transmission power is unable to maintain an acceptable throughput level; the results achieved surpass the findings of [49].

The system illustrated in Figure 5-b has the SNR of 40 dB. At low mobility rates (0 to 30 km/h), throughput is improved, especially with the application of RIS. The RIS enhances signal reception by efficiently reflecting or refracting signals in the desired direction. As mobility increases, the throughput decreases. RIS has no impact on enhanced throughput at high speeds (up to 350 km/h) compared with systems without RIS [51]; the obtained results are better. At a higher SNR level of 40 dBm, throughput always shows that greater transmission of power reduces losses caused by mobility.

attributed to the fact that the RIS system improves signal quality by reducing interference and enhancing channel gain. This is evident in the large changes in the percentage of 1000 users. The results are better than [13-48-50].

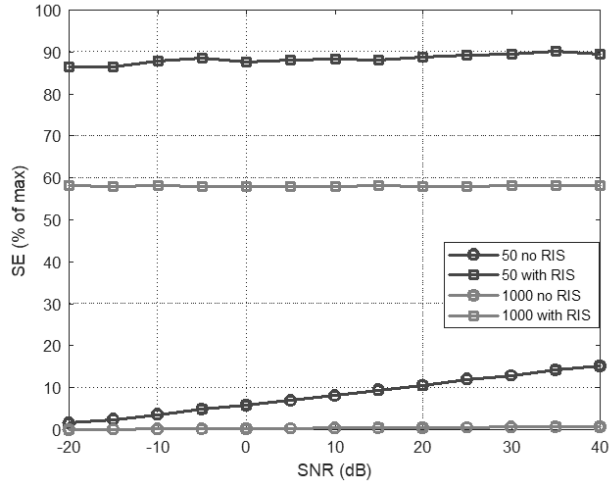


Fig. 3. SE versus SNR for 4 different groups of mMIMO DL CR cooperative NOMA users with and without intelligent RIS

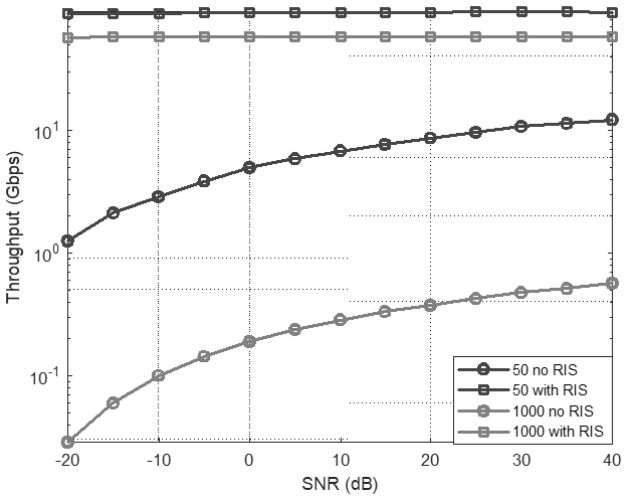


Fig. 4. Throughput vs. SNR for 4 different groups of mMIMO DL CR cooperative NOMA users with and without intelligent RIS

and without intelligent RIS and CR, with the impacts of mobility speed and network density on throughput and latency across different user densities.

Table .1 presents comprehensive details on the simulator settings employed for modeling proposal system networks

Parameter	Value
Number of Users	50 to1000
Mobility Speeds	0 to 350 Km/h
Number of Antennas	256x256
RIS Configuration	512x512
SNR (dB)	-20 to 40
Modulations	1024 QAM
Path-loss exp.	2.7
BW	30 G Hz
Cellular type	Microcells
Frequency Range	26 GHz to 40 GHz

Figure 3 shows the SE and SNR for 50 and 1000 users in the mmWave mMIMO DL CR cooperative NOMA PD system. The SE improved as the SNR rose. The group of 50 and 1000 users reaches a maximum SE of 2.0201 and 0.0952 bps/Hz, respectively, when the combination system is used with the intelligent RIS; the SE is enhanced by 72% and 58%, respectively, at the SNR of 40 dB.

The throughput and SNR for 50 and 1000 users in the mmWave mMIMO DL CR cooperative NOMA PD system are shown in Figure 4. The throughput improved as the SNR increased. The groups of 50 and 1000 users reach a maximum throughput of 12.2143 and 0.5692 Gbps, respectively, when the combination system is used with the intelligent RIS; the throughput is boosted by 76.2% and 98%, respectively, at the SNR of 40 dB.

The RIS system significantly improves the SE and throughput of the proposed system, especially as the number of users increases. This is

Algorithm 1. Q-Learning algorithm

1. Initialize the Q-Table and hyper parameters $(\alpha,\gamma,\varepsilon,\lambda,\mu,\beta)$
2. Commence ε:
   a) Initialize state s, encompassing $SINR, P_{alloc}, and\ P_{Loss}$ R for each user.
3. In the ε of convergence or maximum iterations:
   a) Select action according to the ε-greedy policy.
   b) Implement action (P_{K}) and monitor:
      - Subsequent state s'.
      - Reward $R(s,a)$ is determined by delay, packet loss, and fairness.
   c) Revise Q-value:
      $Q(s,a)\rightarrow Q(s,a)+\alpha * (R(s,a)+\gamma * max Q(s',a')-Q(s,a).$
   d) Update state $s\,to\,s'$.
4. End ε:
   a) Derive the best strategy for latency, packet loss, and equity.
5. Implement an effective power allocation strategy.
6. End.

3. Simulation Results and Discussion

The simulation parameters for the proposal systems model in 6G networks can be seen in Table 1. The results indicate the relationship between the scalability of these systems and the rise in user count, SNR, and throughput. The charts depict the results of many mMIMO cooperative NOMA scenarios, emphasizing the distinctions between configurations with

$$a = \{P_1, P_2, \ldots, P_K\} \tag{30}$$

a represents the power location of each user, determined by ȯ the probability of exploration.

$$R(s,a) = -\frac{1}{K}\sum_{k=1}^{K}\left(L_k + \mu \cdot P_{Loss} R_k\right) + \beta \cdot F \tag{31}$$

Based on the Bellman equation, Q-learning updates Q-values,

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left[R(s,a) + \gamma a' max Q(s',a') - Q(s,a)\right] \tag{33}$$

Q-value for state s and action a; α = learning rate. Discount factor (γ factor). The maximum Q-value for the subsequent state s' over all potential actions s'. The procedure runs until Q values converge or for a set number of iterations. After that, the latency, packet loss and fairness are determined using equations (27, 29 and 30) respectively.

Where L is latency, $f(SNR)$ is the channel-dependent factor for different channel models and user setups.

For each user, packet loss P_{loss} occurs if their SINR falls below a specified SINR threshold,

$$P_{loss} = \begin{cases} 1 & if\ SINR < SINR_{threshold} \\ 0 & otherwise \end{cases} \tag{28}$$

$$F = \frac{\left(\sum_{k=1}^{K} \Sigma r_k\right)^2}{K \cdot \sum_{k=1}^{K} r_k^2} \tag{30}$$

where F is fairness, r_k represents user $k's$ throughput, L_k represents user $k's$ latency, and β balances latency and fairness. The cooperative NOMA throughput, latency, and packet loss optimization approach is thoroughly analyzed, as seen in logarithm 1, focusing on the mathematical aspects of logarithmic power distribution and its effect on latency, packet loss and fairness index calculation. A broad goal is,

$$R = \sum_{k=1}^{K} r_k - \lambda \cdot L - \mu \cdot P \tag{31}$$

where R is System performance reward. r_k is user k throughput. K user total. L system aggregate latency. P System-wide packet loss. The weight factor λ balances throughput and latency. μ Weight factor for packet loss reduction. Each state *s* encapsulates pertinent attributes of the system:

$$s = \{SINR_1, SINR_2, \ldots, SINR_K,\ P_{alloc},\ PLR_1, PLR_2, \ldots, PLR_K\} \tag{32}$$

where power allocation P_{alloc}, user k (P_{Loss} R) packet loss ratio. Power allocation to each user is action as,

$$p_{\mathrm{k}}(t+\Delta t)=p_{\mathrm{k}}(t)+v_{\mathrm{k}}\Delta t\cdot\left[cos(\theta_{\mathrm{k}}),sin(\theta_{\mathrm{k}})\right] \tag{18}$$

Mobility speed in m/s is represented by $vi \in MobilitySpeeds$. Modeling the BS-user channel uses path loss, small-scale fading, and user mobility. For user k at distance $d_{\mathrm{k}}(t)$, the path loss is [48],

$$PL_{\mathrm{k}}(t)=\left(\frac{4\pi d_{\mathrm{k}}(t)}{\lambda}\right)^{-\alpha} \tag{19}$$

All users' small-scale fading is modelled as Rayleigh fading channels. At time t , the BS-user k channel matrix $H_{\mathrm{k}}(t)$ is,

$$H_{\mathrm{k}}(t)=\sqrt{PL_{\mathrm{k}}(t)}\cdot\left(G_{\mathrm{k}}(t)+jB_{\mathrm{k}}(t)\right) \tag{22}$$

The real and imaginary channel sections are modelled by independent and identically distributed Gaussian random matrices $G_{\mathrm{k}}(t)$ and $B_{\mathrm{k}}(t)$. RIS channels connect users to the RIS. $H_{RIS,k}(t)$ represents the channel matrix for RIS-assisted user k at time t [49],

$$H_{RIS,k}(t)=\sqrt{PL_{RIS,k}(t)}\cdot\Phi(t)\cdot\left(G_{RIS,k}(t)+jB_{RIS,k}(t)\right) \tag{23}$$

The real and imaginary components of the channel from RIS to user k are $G_{RIS,k}(t)$ and $jB_{RIS,k}(t)$.

$$SINR_{RIS,k}(t)=\frac{P_k(t)\cdot\left|H_{RIS,k}(t)\right|^2}{\sum_{j<k}P_j(t)\cdot\left|H_{RIS,k}(t)\right|^2+N_0} \tag{25}$$

$$D=SE\times BW \tag{26}$$

where D is data rate (bps).

$$L=\frac{Data\ Size}{D}\times f(SNR) \tag{27}$$

where I is user interference. RIS signal intensity is greatly improved by optimizing phase shifts. Define optimization issue:

$$\max_{\theta} \left| H_{RIS} \right|^2 \tag{13}$$

Combining all components creates this cooperative NOMA system with RIS and CR for SE. Both direct and RIS-assisted user-BS connections employ the Rayleigh fading channel model. The user $k's$ effective NOMA system with RIS channel is,

$$h_{eff,\mathrm{k}} = H_{\mathrm{k}} + \mathrm{g}_{\mathrm{k}}^{H} \Theta H_{d} \tag{14}$$

The SE for user k in a NOMA system with RIS [46] is,

$$SE_{\mathrm{k}} = \log_2 \left(1 + \frac{P_{\mathrm{k}} \left| h_{eff,\mathrm{k}} \right|^2}{\sum_{i<\mathrm{k}} P_i \left| h_{eff,i} \right|^2 + \sigma^2} \right) \tag{15}$$

Calculate throughput T as,

$$T = SE_{total} \times B_{avail} \tag{16}$$

To calculate throughput with and without RIS, alter user density by changing the number of users k depending on area size A and density ρ, $K = \rho \times A$. Thus, user density affects throughput [47],

$$T_{with/without\ RIS}(\rho) = \sum_{k=1}^{\rho \times A} SE_{with/without\ RIS,k} \tag{17}$$

Changes in distance and relative velocities alter channel conditions when users move randomly. Let, $p_{\mathrm{k}}(t) = \left[x_{\mathrm{k}}(t), \mathcal{Y}_{\mathrm{k}}(t) \right]$ represent user $k's$ position at time t. v_{k} represents user iii's speed. θ_{k} represents user $k's$ random movement direction. The position of user k at time $t + \Delta t$ is,

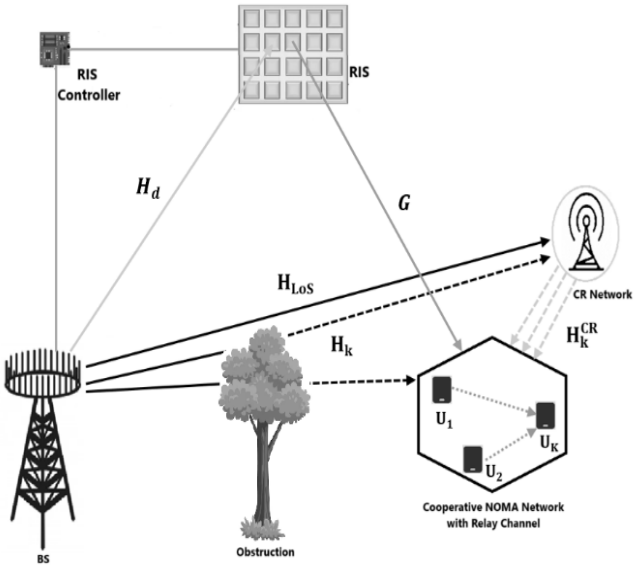


Fig. 2 illustrates mMIMO PD DL CR mMIMO cooperative NOMA with users employing intelligent IRS and mm-Wave methodologies

The RIS-assisted communication connection CSI channel model is,

$$H_{RIS} = G\Theta H_d \tag{10}$$

The BS-RIS channel is represented by $H_d \in \mathbb{C}^{N_{RIS} \times M}$, whereas the RIS-user channel is $G \in \mathbb{C}^{N \times N_{RIS}}$. The diagonal matrix $\Theta = diag(e^{j\theta_1}, e^{j\theta_2}, \ldots, e^{j\theta_{N_{RIS}}})$ represents the phase shifts caused by the RIS elements. The user received this RIS signal [42],

$$\mathcal{Y}_k^{RIS} = H_k^H x + g_k^H \Theta H_d x + z_k \tag{11}$$

where g_k is the RIS-user channel.

$$SINR_{RIS} = \frac{P_k \cdot |H_{RIS}|^2}{I + N_0} \tag{12}$$

User index k (1st user has most power). In cooperative NOMA, user $k's$ received signal is,

$$y_k = H_k^H \sum_{\mathrm{n}=1}^{K} \sqrt{P_{\mathrm{k}}}\, x_{\mathrm{k}} + z_k \tag{6}$$

x_{k} is the transmitted signal, P_{k} is the power (higher for poorer channel conditions), and z_k is noise at user k. SIC lets better channel users decode and delete weaker channel users' signals before decoding their own. The SINR for k users as [11],

$$SINR_k = \frac{P_k \cdot h_k}{\sum_{j=1, j\neq k}^{K} P_j \cdot h_j + N_0} \tag{7}$$

For cooperative users, SINR additionally accounts for relayed signals [44],

$$SINR_{coop} = SINR_k + \sum_{j=1}^{k-1} SINR_j \tag{8}$$

Calculate the k^{th} user's SE_k,

$$SE_k = \frac{B}{K} \log_2 \left(1 + SINR_{coop}\right) \tag{9}$$

The diverse reflective elements of the intelligent RIS device can alter the phase of incoming electromagnetic waves, hence improving communication quality, as seen in Figure 2.

distribution is employed to determine the real and imaginary channel coefficients for Rayleigh fading.

$$y_{\mathrm{k}} = H_{\mathrm{k}}^{H} x + z_{\mathrm{k}} \tag{2}$$

where x is the BS broadcast signal, $z_{\mathrm{k}} \sim C\mathcal{N}\left(0, \sigma^2\right)$ is the AWGN with variance σ^2. CR comprises PUs, who are authorized to utilize certain frequency bands, and SUs, who can access unoccupied spectrum without disrupting the PUs. Spectrum sensing methodologies allow PUs to detect available channels. A power discovery technique is employed to assess channel occupancy by monitoring power levels, while the presence of primary users is identified using prior knowledge of their signal characteristics, mathematically described as follows.

$$E = \sum_{n=1}^{N} \left| \mathcal{Y}[n] \right|^2 \tag{3}$$

The receive signal $\mathcal{Y}[n]$, N samples. The spectrum sensing decision rule is [44],

$$\text{Decision} = \begin{cases} \mathrm{H}_0 & \text{if } E < \lambda \left(\text{spectrum available}\right) \\ \mathrm{H}_1 & \text{if } E \geq \lambda \left(\text{spectrum occupied}\right) \end{cases}$$

E is the signal energy and λ is the detection threshold. The available bandwidth, B_{avail}, is determined as follows,

$$B_{avail} = B_{total} - B_{occupied} \tag{4}$$

$B_{occupied}$ represents the bandwidth utilized by PUs ascertained using spectrum sensing. NOMA multiplexes PD users to share time-frequency resources. Channel conditions determine user power levels. The total power Pt is allocated among users based on their channel conditions.

$$P_k = P_t \cdot \frac{1}{k} \tag{5}$$

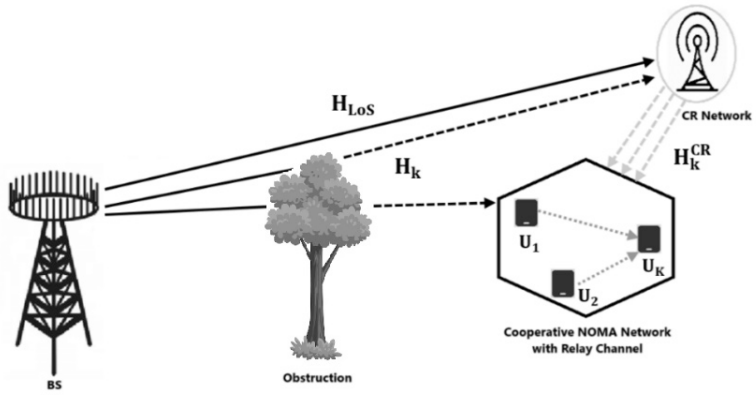


Fig.1 illustrates mMIMO PD DL mMIMO cooperative NOMA with k users employing CR and mmWave technology

In mMIMO, each BS possesses several antennas (M) and accommodates multiple users (N). The mMIMO will physically multiplex many users and focus energy in narrow beams on each user, enhancing data rates and spectrum efficiency. The mMIMO channel is represented for each user as a Rayleigh fading channel matrix. When there is no LoS path between the transmitter and receiver, Rayleigh fading happens. This makes the received signal made up of several separate, scattered parts. Modeling the BS (with M antennas) for user k channel, the $H_{\mathrm{k}} \in \mathbb{C}^{M\times 1}$ channel vector connecting the BS to the $k-th$ user [42].

$$H_{\mathrm{k}}=\left[h_{\mathrm{k,l}}, h_{\mathrm{k},2}, \ldots, h_{\mathrm{k},M}\right]^{T} \tag{1}$$

where is the system user index. $h_{\mathrm{k},M}$ is the channel coefficient between the kth user and the transmit antenna. M_T is the number of BS or transmitter transmitting antennas. The coefficient of the Rayleigh fading channel between the $m-th$ antenna at the BS and the $k-th$ user is denoted by the complex Gaussian random variable $h_{k,m} \sim C\mathcal{N}\left(0, \beta\mathrm{k}\right)$. User k perceives $\beta\mathrm{k}$ as large-scale route loss and shadowing. A formalizable univariate Gaussian

The impact of different mobility speeds on the effectiveness of CR mMIMO DL cooperative NOMA networks, ranging from 0 to 350 km/h, is analysed, considering scenarios with and without RIS. This element is crucial for understanding the robustness of intelligent RIS-assisted systems for high-mobility users.

The examination of user density and the impact of latency, packet loss, and SNR on fairness index throughput demonstrates the scalability of NOMA collaborative DL systems and suggests improvements to the network architecture.

This work employs a Q-learning algorithm to enhance power allocation, providing an adaptive method that enables the system to learn and modify power distribution according to network circumstances, hence optimizing latency, packet loss, and fairness index.

The succeeding portions of this work conform to this structure. Section II offers a detailed examination of the proposed models and approaches. Section III outlines the simulation parameters and reports the results. Section IV delineates the outcomes of these endeavours.

2. System Model

This study investigates wireless networks. The network comprises several mMIMO DL cooperative NOMA user groups that employ CR integration and mm-Wave technology as shown in Figure 1. Users are positioned at varying distances from the BS, leading to varied received power levels. The network utilizes 512-quadrature amplitude modulation (QAM). Number of users and the operating principles of cooperative NOMA inside the power domain (PD).

performance of the NOMA-enhanced network [36]. To promote the use of NOMA systems among users with equivalent transmission power, the creators of [37] improved uplink communication by the incorporation of active and passive RIS. The objective of formulating a hybrid user clustering and RIS allocation approach was to improve the implementation of the NOMA scheme and optimize the system's aggregate rate [38]. The effectiveness of the RIS-enhanced NOMA network was analysed in [39], concentrating on energy efficiency in both delay-tolerant and delay-constrained modes. The authors of [40] developed a deep learning framework and assessed the RIS-assisted CR-NOMA system to forecast ergodic performance.

Currently, most research is on beamforming design employing RIS. Nevertheless, there are several peculiarities regarding our job. While no CR-NOMA network system model [41] presently exists, passive beamforming on RIS is the preferred approach after some simplifications. Therefore, the optimization procedures differ from our work. Thus, due to the characteristic optimization problem presented in [42], even when using analog conventions. The method examined in [43] differs significantly from ours in that we aim to optimize SE and throughput using the CR cooperative NOMA mMIMO network with DL assisted by intelligent RIS. The principal contributions encompass

This study introduces a unique mMIMO DL cooperative NOMA system for many users inside a 6G mmWave communication framework and utilizes clever intelligent RIS and CR. To demonstrate its scalability and adaptability in user-dense contexts, we evaluate it with different user counts (50 and 1000).

SNR, the impact of resource allocation is negligible. The authors performed a comprehensive examination of the efficacy of entire transmit power systems with STAR-RIS [27-28]. The evaluation of performance was conducted concerning both fault-free and non-fault-free cascade interference cancellation, as indicated in [29]. A comparison between NOMA and OMA communication systems was conducted in [30] within the context of phase-shifted coupled STAR-RIS.

The notion of RIS systems has become significant as an economical option for 6G wireless networks. Numerous cooperative NOMA configurations utilizing RIS have been mentioned in the literature. The downlink (DL) transmissions of a system model comprising one BS, two users, and one receiver-input-signal were analyzed in the work referenced as [31]. The concept of utilizing cooperative NOMA resulted in a reduction of the overall transmitted power. The study in [32] examined the application of RIS to enhance the efficiency of cell border users in a SWIPT NOMA system, where data is sent from a user at the cell core to a user at the cell boundary. In [33], researchers examined a two-stage RIS-assisted transmission approach for cooperative NOMA networks with SWIPT, which may enhance the attainable rate for strong users while maintaining the service quality requirements of weak users.

The future of wireless networks is characterized by CR NOMA. Multiple networks can share a single frequency due to CR's sophisticated monitoring and decision-making, enhancing spectrum utilization [34]. NOMA enhances connection, equity, and SE by allowing many users to share time, code, and frequency resources [35]. The ergodic capacity and outage probability (OP) were assessed from the fundamental critical route to evaluate the

communication is unattainable with traditional RIS systems that rely solely on reflection. A proposed solution to this issue, as shown in [21], is STAR-RIS, which denotes simultaneous transmission and reflection employing RIS.

Recent research on mMIMO-NOMA in mmWave/THz networks has predominantly overlooked user clustering in favor of performance analysis. To meet the increased expectations for spectral efficiency and multiple user connections in 6G, NOMA-enabled networks must also include organizational user grouping. Moreover, user clustering in networks functioning in low-frequency bands has garnered considerable study attention, but mmWave/THz networks remain largely unexamined. Nonetheless, investigations on user pairing in the context of clustering users within a MIMO-NOMA system are limited to a restricted number of users [22-23]. Recent studies classify users as cellular or device-to-device (D2D) via a cluster-matching technique grounded in channel correlation [24].

This method converts user clustering into a polynomial problem. Despite its general simplicity, a hurdle in learning-assisted clustering systems is the insufficient initialization of cluster heads. The efficacy of the STAR-RIS-based system for simultaneous transmission and reflection in fading channels has been previously examined by a compendium of authors [25-26]. The findings indicate that STAR-RIS surpasses conventional RIS in NOMA systems, especially for users located at the cell periphery lacking a direct line of sight (LoS) to the base station (BS). The authors illustrated that in low signal-to-noise ratio (SNR) scenarios, the inverse disproportionate allocation of resources in STAR-RIS can offset the BS's uneven power distribution, establishing that NOMA utilizing STAR-RIS remains effective despite the absence of a direct link between the BS and the users. In areas with a high

Various multiple-access algorithms exist, including non-orthogonal multiple access (NOMA). As seen in [7-8], NOMA can increase spectral efficiency (SE) and user throughput service. Receivers in mobile devices reduce beam-induced interference by the application of successive interference cancellation (SIC) techniques. According to [9], the fundamental concept of NOMA allows for the integration of many users by categorizing them based on cost or time. The greater the number of individuals utilizing NOMA, the more orthogonal resources become available [10-11]. The advancement of NOMA is anticipated to address these difficulties. These systems efficiently accommodate a substantial number of users owing to their optimized architecture and reduced resource consumption [12]. A compelling option that demonstrates the potential for enhancing data throughput in mmWave and THz communication systems is massive multiple-input multiple-output (mMIMO) NOMA [13-14].

Cognitive radio (CR) technology is an effective method for spectrum management, providing opportunistic, on-demand connections that resolve capacity challenges in traditional licensed wireless communication networks. Primarily, in CR-based networks, there are PUs, or authorized users, and SUs, or unauthorized users. The SUs aims to exploit any opportunities that arise while they contend with the PUs for licensed spectrum [15-16]. Consequently, CR technology will be essential in the advancement of future wireless networks and in meeting their exceedingly quick demands [17].

An approach to creating intelligent and efficient radio configurations is the idea of reconfigurable intelligent surfaces (RIS) [18- 19-20]. A compilation of passive reflecting elements (RE) forms RIS, which systematically alter the phase and amplitude of incoming signals. However, comprehensive spatial

1. Introduction

Every generation of wireless networks, from first-generation (1G) to sixth-generation (6G), has seen enhancements in capacity and quality of service [1-3]. Establishing a 6G network necessitates surmounting several technological challenges, such as the implementation of intricate coding methodologies, enhanced frequency bands, and the development of novel antenna technologies. The 6G wireless network aims to achieve several goals, including reduced latency, increased throughput, extensive connection, and enhanced energy and spectrum economy. The volume of data transmission has increased markedly recently due to the widespread use of intelligent devices and equipment. Significant advancements have been made in the terahertz (THz) and millimeter-wave (mmWave) domains to meet the anticipated high demand [4-6]. To commence the 6G network, it is important to tackle technological problems such as the improvement of coding procedures, the optimization of frequency bands, and the advancement of antenna technology. 6G technology is expected to transform several industries upon its implementation due to its ability to provide enhanced degrees of immersion and engagement. The included industries include healthcare, transportation, and entertainment. Future communication networks are expected to demonstrate a substantial demand for 6G wireless technology. Wireless devices and services, including VR/AR, the Internet of Everything, and virtual reality, are seeing increased popularity and widespread acceptance, hence presenting new obstacles to create wireless networks of the future. The demand for varied data transfer, increased connection density, reduced latency, and optimal bandwidth utilization intensifies these challenges.

Abstract

This study examines spectral efficiency (SE) and throughput throughout a spectrum of user densities (from 50 to 1000 users), user mobility speeds (0 to 350 km/h), latency, packet loss, and fairness index, within a wide range of signal-to-noise ratios (SNRs). The analysis includes a number of different situations, such as (i) cooperative non-orthogonal multiple access (NOMA) with massive multiple-input multiple-output (mMIMO), (ii) mMIMO cooperative NOMA integrated with cognitive radio (CR), and (iii) CR-assisted mMIMO cooperative NOMA enhanced with reconfigurable intelligent surfaces (RIS). All of these are part of 6G millimeter-wave (mmWave) networks. The study investigates the enhancement of latency, packet loss, and fairness index in proposed systems by a unique approach that dynamically optimizes power distribution via a Q-learning algorithm. The mathematical clarification of each equation provides a comprehensive understanding of signal reception by users, the dynamics and implications of CR, and the influence of intelligent RIS optimization on it. The results show that adding the IRS improves resource allocation, makes users perform better in crowded areas, lowers latency and packet loss, and raises the fairness index by reducing interference and making channel access more efficient, especially when using the suggested optimization algorithm. The results help 6G technology move further with scalable and efficient communication networks.

Keywords: cooperative NOMA, Massive MIMO, CR, reconfigurable intelligent surfaces (RIS), SE, millimeter wave (mmWave).

Dynamic RIS-Enabled Massive MIMO NOMA Systems Power Allocation Optimization for 6G Using Machin learning Approach/ Original article

Mohamed Hassan[1], Khalid Hamid[2], Salah Hagahmoodi[3], Elmuntaser Hassan[4],

1,2 Department of Electrical Engineering, Omdurman Islamic University, Omdurman, Sudan

1 mhbe4321@gmail.com, 2 khalid.bilal@oiu.edu.sd

3 Salah Hagahmoodi, Department of Information Technology, College of Computer Science and Information Technology, Almoughtarbeen University, Sudan.

salahhagahmoodi@gmail.com mailto:salahhagahmoodi@gmail.com

4 Department of Computer Science, Al-Neelain University, Khartoum, Sudan

3 Habba86@hotmail.com

Corresponding Author: Salah Hagahmoodi, salahhagahmoodi@gmail.com

[44]. Rhoades, J. D., Kandiah, A., & Mashali, A. M. (1992). *The Use of Saline Waters for Crop Production* (FAO Irrigation and Drainage Paper 48). Food and Agriculture Organization of the United Nations. https://www.fao.org/3/t0667e/t0667e00.htm

[45]. Rhoades, J. D., Kandiah, A., & Mashali, A. M. (1992). *The use of Saline Waters for Crop Production* (FAO Irrigation and Drainage Paper 48). Food and Agriculture Organization of the United Nations. https://www.fao.org/3/t0667e/t0667e00.htm

[46]. Richards, L. A. (1954). *Diagnosis and Improvement of Saline and Alkali Soils* (Issue 60). US Government Printing Office.

[47]. Richards, L. A. (1954). *Diagnosis and Improvement of Saline and Alkali Soils* (Issue 60). US Government Printing Office.

[48]. Rivett, M. O., Buss, S. R., Morgan, P., Smith, J. W., & Bemment, C. D. (2008). Nitrate Attenuation in Groundwater: A review of Biogeochemical Controlling Processes. *Water Research, 42*(16), 4215–4232. https://doi.org/10.1016/j.watres.2008.07.020

[49]. Saeed, T. (2014). Assessment and Conservation of Groundwater Quality: A Challenge for Agriculture. *British Journal of Applied Science and Technology*. https://doi.org/10.9734/BJAST/2014/6353

[50]. Srivastav, A. L., Dhyani, R., Ranjan, M., Madhav, S., & Sillanpää, M. (2021). Climate-resilient Strategies for Sustainable Management of Water Resources and Agriculture. *Environmental Science and Pollution Research*. https://doi.org/10.1007/S11356-021-14332-4

[51]. U.S. Salinity Laboratory Staff. (1954). *Diagnosis and Improvement of Saline and Alkali Soils*. USDA Agricultural Handbook No. 60.

[52]. UNESCO. (2021). *World Water Development Report 2021: Valuing Water*. United Nations Educational, Scientific and Cultural Organization.

[53]. USDA-NRCS. (1997). *National Engineering Handbook, Part 652 – Irrigation Guide*. United States Department of Agriculture – Natural Resources Conservation Service.

[54]. Varela-Ortega, C., Blanco-Gutiérrez, I., Swartz, C. H., & Downing, T. E. (2011). Balancing Groundwater Conservation and Rural Livelihoods Under Water and Climate Uncertainties: An Integrated Hydro-economic Modeling Framework. *Global Environmental Change-Human and Policy Dimensions*. https://doi.org/10.1016/J.GLOENVCHA.2010.12.001

[55]. White, P. J., & Broadley, M. R. (2003). Calcium in Plants. *Annals of Botany*, 92(4), 487–511. https://doi.org/10.1093/aob/mcg164

[56]. World Health Organization (2017). *Guidelines for Drinking-water Quality* (4th ed.). WHO Press

[28]. Jassim, M., & Al-Ani, K. (2022). Hydrochemical Characteristics and Irrigation Suitability of Groundwater in Balad, Iraq. *Agricultural Water Management*, 30(1), 112-125

[29]. Kareem, A., Saeed, H., & Jaber, M. (2019). Microbiological Contamination of Groundwater in Samarra: Implications for Public Health. *Environmental Health Perspectives*, 27(3), 150-165.

[30]. Kareem, A., Saeed, H., & Jaber, M. (2020). Assessment of Potable Water Quality in Balad city, Iraq. *Environmental Research Letters*, 15(5), 205-220.

[31]. Kareem, M. A., Al-Mashhadani, M. S., & Hussein, K. J. (2011). Groundwater Management in Northern Iraq. *Environmental Earth Sciences*, 62(6), 1193–1204. https://doi.org/10.1007/s12665-010-0604-y

[32]. Li, C., Li, C., Gao, X., Li, S., & Bundschuh, J. (2020). A review of the Distribution, Sources, Genesis, and Environmental Concerns of Salinity in Groundwater. *Environmental Science and Pollution Research*. https://doi.org/10.1007/S11356-020-10354-6

[33]. Lindsay, W. L. (1979). *Chemical Equilibria in Soils*. New York: Wiley.

[34]. Loneragan, G. H., Gould, D. H., Callan, R. J., Sigurdson, C. J., & Mason, G. L. (2001). Association of Excess Sulfur Intake and An increase in Cerebrocortical Necrosis in Feedlot Cattle. *Journal of the American Veterinary Medical Association, 219*(8), 1167–1174. https://doi.org/10.2460/javma.2001.219.1167

[35]. Mahdi, A. A., & Al-Jaberi, R. T. (2021). Impact of Human Activities on Groundwater Quality in Central Iraq. *Iraqi Journal of Science*, 62(4), 1003–1015.

[36]. Mahmood, S. A., & Jassim, A. M. (2023). Groundwater Quality Assessment for Irrigation in Iraq. Water Resources Management, 37(4), 765-780.

[37]. Mahmoud, S., & Hassan, N. (2021). Heavy Metal Contamination in Groundwater Sources in Salah Al-Din Governorate. *Iraqi Journal of Water Resources*, 18(2), 75-89.

[38]. Marschner, H. (2012). *Mineral Nutrition of Higher Plants*. Academic Press.

[39]. Meharg, A. A., & Rahman, M. M. (2003). Arsenic Contamination of Bangladesh Paddy Field Soils: Implications for Rice Contribution to Arsenic Consumption. *Environmental Science & Technology*, 37(2), 229–234. https://doi.org/10.1021/es0259842

[40]. Minhas, P. S., & Tyagi, N. K. (1998). Guidelines for Irrigation with Saline and Alkali Waters. *Bulletin No. 1/98*, CSSRI, Karnal, India.

[41]. Ministry Of Agriculture of Iraq, Agroecological Zoon Department (AEZ).

[42]. Munns, R., & Tester, M. (2008). Mechanisms of Salinity Tolerance. *Annual Review of Plant Biology*, 59, 651-681.

[43]. Ramesh, K., & Elango, L. (2012). Groundwater Quality Assessment for Irrigation Use in Vellore District, Southern India. *Environmental Earth Sciences*, 67(8), 2129–2141. https://doi.org/10.1007/s12665-012-1656-9

[14]. Ayers, R. S., & Westcot, D. W. (1985). *Water Quality for Agriculture*. FAO Irrigation and Drainage Paper No. 29 Rev. 1, Food and Agriculture Organization of the United Nations. FAO Link

[15]. Ayers, R. S., & Westcot, D. W. (1985). *Water Quality for Agriculture*. FAO Irrigation and Drainage Paper 29 Rev. 1. Rome: Food and Agriculture Organization.

[16]. Cakmak, I. (2008). Enrichment of Cereal Grains with Zinc: Agronomic or Genetic Biofortification? *Plant and Soil*, 302(1-2), 1-17.

[17]. Ehab Mohammad, A., Ektifa Taha, A., Fatma Adnan, S., & Rania Haithem, S. (2020). Study of qualitative properties of groundwater and its suitability for different uses in the Eastern of the Al-Dour city/Salahaldin/Iraq. *Tikrit Journal of Pure Science*, *25*(2), 47-53.https://doi.org/10.25130/J.V25I2.957

[18]. FAO. (2019). *Iraq – Water Report 37*. Food and Agriculture Organization of the United Nations.

[19]. Grattan, S. R., & Grieve, C. M. (1999). Salinity–mineral nutrient relations in horticultural crops. *Scientia Horticulturae*, 78(1-4), 127–157. https://doi.org/10.1016/S0304-4238(98)00192-7

[20]. Gupta, U. C., & Gupta, S. C. (1987). Trace Element Toxicity Relationships to Crop Production and Livestock and Human Health: Implications for Management. *Communications in Soil Science and Plant Analysis, 18*(8), 931–970. https://doi.org/10.1080/00103628709367847

[21]. Hamza, S., & Younis, T. (2021). Assessment of Heavy Metals in Groundwater Sources in Samarra. *Iraqi Journal of Environmental Studies*, 19(2), 89-104.

[22]. Hawkesford, M. J. (2000). Plant Responses to Sulphur Deficiency and the Genetic Manipulation of Sulphate Transporters to Improve S-utilization Efficiency. *Journal of Experimental Botany, 51*(342), 131–138. https://doi.org/10.1093/jexbot/51.342.131

[23]. Horneck, D. A., Ellsworth, J. W., Hopkins, B. G., Sullivan, D. M., & Stevens, R. G. (2007). Managing salt-affected soils for crop production. *Pacific Northwest Extension Publication PNW601*. Oregon State University. https://catalog.extension.oregonstate.edu/pnw601

[24]. Hussain, A. J., *et al*. (2022). Evaluation of Groundwater Salinity in Central Iraq. *Environmental Earth Sciences*, 82(5), 471.

[25]. Ismail, A.H., Shareef, M.A., Hassan, G., & Alatar, F.M. (2023). Hydrochemistry and Water Quality of Shallow Groundwater in the Tikrit Area of Salah Al Din Province, Iraq. *Applied Water Science*, 13, Article 197. https://doi.org/10.1007/s13201-023-02008-y

[26]. Jabbar, A. J. (2022). Water Quality Index and Heavy Metals Evaluation in Al-Dour Water Treatment Plant. *Indian Journal of Environmental Protection*, 42(4), 330–336. https://www.indianjournals.com/ijor.aspx?article=031&issue=4&target=ijor%3Aije1&volume=48

[27]. Jassim, M., & Al-Ani, K. (2022). Evaluation of Groundwater Quality for Irrigation in Samarra, Iraq. *Agricultural Water Management*, 31(2), 125-140

References

[1]. Abd Al Satar, N. H., & Sachit, D. E. (2021). The effect of Hospital Wastewater Discharge of Medical City, Baghdad on Heavy Metals Concentration of the Tigris River. *Desalination and Water Treatment*, 230, 252–258.

[2]. Abd Al Satar, Nawras H, and Dawood E Sachit. 2021. "Assessment of Hospital Wastewater Quality and Management in Bab-Al Muadham Region at Baghdad." *Journal of Engineering and Sustainable Development* 25(3): 44–50.

[3]. Abed, M.F., Zarraq, G., & Ahmed, S.H. (2021). Hydrogeochemical Assessment of Groundwater Quality and its Suitability for Irrigation and Domestic Purposes in Rural Areas, North of Baiji City-Iraq. *Iraqi Journal of Science*, 62(7), 2296–2306. https://doi.org/10.24996/ijs.2021.62.7.18

[4]. Acharya, G. D., Solanki, M. R., & Hathi, M. V. (2010). Studies on Physico-Chemical Parameters of Irrigation Water, Prantij, Gujarat (India). *International Journal of Chemical Sciences*.

[5]. Addiscott, T. M. (2005). Nitrate, Agriculture and the Environment. *Agriculture, Ecosystems & Environment, 109*(1–2), 1–2. https://doi.org/10.1016/j.agee.2005.02.001

[6]. Ajmal, P. M., Babar, M., Gul, N., & Jokhio, R. (2023). Assessment of Groundwater Quality Using Water Quality Index, and Geo-Spatial Tools. *Neutron: Jurnal Rekayasa Teknik Sipil*. https://doi.org/10.29138/neutron.v22i2.180

[7]. Al-Ansari, N., Knutsson, S., & Aljawad, S. (2018). Water Scarcity in Iraq: Challenges and Solutions. *Engineering*, 10(3), 59–71.

[8]. Al-Dabbas, M. A., Al-Ansari, N., & Knutsson, S. (2018). Hydrogeochemical Assessment of Groundwater in Northern Iraq. *Environmental Earth Sciences*, 77(6), 256.

[9]. Ali, M. A., *et al*. (2022). Microbial Contamination of Groundwater in Al-Alam District. *Revista Bionatura*, 7(1), 26.

[10]. Ali, S. H., & Al-Haidari, N. A. (2021). Environmental Risk Assessment of Industrial Activities on Groundwater in Baiji, Iraq. *Journal of Environmental Science and Pollution Research*, 28(18), 23052–23061.

[11]. Al-Jiburi, H. K., & Al-Basrawi, N. H. (2021). Hydrogeological Assessment of Groundwater in Salah Al-Din Governorate. *Iraqi Journal of Science*, 59(3), 145-160

[12]. Al-Saadi, R., Mohammed, A., & Younis, T. (2019). Evaluation of Groundwater Quality in Balad District, Iraq. *Journal of Environmental Studies*, 12(3), 45-60.

[13]. Al-Tameemi, R., Mohammed, A., & Saleh, T. (2020). Hydrochemical Analysis of Groundwater in Samarra, Iraq. *Journal of Water Research*, 22(1), 67-80.

Conclusion

Based on this study, it was found that it is essential to conduct a comprehensive analysis of the physical and chemical properties of groundwater in Salah Al-Din before using it for irrigation, due to its significant impact on soil and agricultural productivity. It is also important to assess the suitability of well water for various types of crops in order to ensure effective agricultural planning, better utilization of arable land, and higher yields. The results indicated that the electrical conductivity of water plays a significant role in the growth and productivity of many agricultural crops. High salinity levels were observed in the groundwater of Salah Al-Din Governorate, exceeding the permissible limits for irrigation water, in addition to elevated concentrations of heavy metals. Therefore, the study recommends the treatment of groundwater before its use in irrigation, to avoid any potential risks to the environment, plants, and consequently, human health.

The following table is for one of the wells in the studied area which shows the effect of EC on crop yield where EC value was 7.42:

(Table1: salinity effect on plants and crop yield for location 1)

Crop	ECen	ECmax	ECe	Yield
Barley	8	28	7.42	100
Cotton	7.7	26.9	7.42	100
Sugar beet	7	24	7.42	97.522
Date palms	4	31.8	7.42	87.688
Wheat	6	20.1	7.42	89.918
Maize	1.7	10	7.42	31.36
Potato	1.7	10	7.42	31.36
Beas	1	6.3	7.42	21.98-
Onions	1.2	7.5	7.42	0.48
Rice	3	11.3	7.42	46.96
Citrus (Orange)	1.7	8	7.42	29.28
Groundnut	3.2	6.6	7.42	65.88-
Carrots	1	8.1	7.42	10.12
Apricot	1.6	5.8	7.42	39.68-
Peas	1.5	11.3	7.42	17.12
Lettuce	1.5	9.8	7.42	28.96
Broccoli	2.8	13.7	7.42	57.496
Grapes	1.5	12	7.42	43.168
Alfalfa	2	15.7	7.42	60.434
Sugar cane	1.7	18.6	7.42	66.252
Clover Barseem	1.5	19	7.42	66.256
Sorghum	6.8	13.1	7.42	90.018
Soybean	5	10	7.42	51.6
Cowpeas	1.5	13.2	7.42	69.76
Spinach	2.6	12.2	7.42	42.642
Beet red	4	15.1	7.42	69.22

lead to reduced yields, soil degradation, and long-term environmental issues. Understanding the relationship between groundwater quality and crop suitability is essential for sustainable agriculture and food security. The following studies are showing the effect of groundwater quality on crops:

- Rhoades *et al*. (1992) found that long-term irrigation with saline water leads to accumulation of salts in the root zone, damaging salt-sensitive crops and restricting options to only salt-tolerant varieties like barley or cotton.
- Richards (1954) noted that water with high SAR levels causes dispersion of soil particles, leading to poor soil tilth and adverse effects on most field crops.
- Ayers and Westcot (1985) highlighted that irrigation water with pH outside the optimal range could interfere with nutrient uptake, particularly affecting crops like maize and rice.
- In parts of Bangladesh and India, arsenic-contaminated groundwater used for irrigation has led to bioaccumulation in rice, posing a serious food safety issue (Meharg & Rahman, 2003).
- A study in Rajasthan, India, showed that in saline groundwater areas, farmers replaced traditional wheat and chickpea with salt-tolerant mustard and barley to cope with declining water quality (Minhas & Tyagi, 1998).
- A study by Ministry of Agriculture in Iraq was done to explain the effect of EC on many crops in Samarra in Salah Al-Din Governorate, Iraq using data of several wells and using the following formula to calculate yield for crops depending on EC values: **Yield=100 - ECemax (ECe - ECen)**

to use for personal remedies against diseases (Kareem *et al.*, 2019). Research work carried out by Jassim & Al-Ani (2022) assessed the quality of irrigation water indices and observed that more than 50% of the tested samples possessed high salinity and SAR and were non-suitable for direct agricultural utilization. Soil management and dilution are also required to reduce these effects (Jassim & Al-Ani, 2022). Several indices such as SAR (Sodium Adsorption Ratio), Na% (Sodium Percentage) and RSC (Residual Sodium Carbonate) have been used to evaluate the efficiency of groundwater for irrigation. It is pointed out that most of the Tikrit soil samples were found to be in "non-suitable" category as a result of severe salinity and sodicity threats (Abed *et al.*,2021). aquifers of Al-Dour indicate high levels of salinity, hardness, and mixture with some heavy metals. These are problems that are not only harmful to human health, but which damage agricultural productivity. Local water treatment such as reverse osmosis should be adopted, awareness of water quality should be raised among the general public, and periodic programs for ground water monitoring should be established (Kareem *et al.*, 2011; WHO,2017). The pH of groundwater samples ranged from 6.4 to 8.4, falling within WHO standards. However, TDS values from 1380 to 1903 mg/L and total hardness from 1843 to 2357 mg/L indicate that the water is very hard and mineralized. Sulfate concentrations exceeded 2000 mg/L in some samples, far beyond the WHO limit of 250 mg/L. Additionally, heavy metals such as iron and cadmium were detected above permissible levels, potentially from corrosion of distribution pipes or geochemical sources (Jabbar, 2022).

4. The Effect of Groundwater Quality on Crop Types

The quality of groundwater directly influences agricultural productivity and the types of crops that can be successfully cultivated. Poor-quality groundwater contaminated with salts, heavy metals, or toxic chemicals can

this effect. (2019) found in the studied soils electrical conductivity in the range 590–3492 µS/cm, which indicated the different levels of soil solution salts. The pH values were between 7.02 and 7.85, that were close to the neutral. Also, it was noted that sulfate concentration had a wide range of difference from (49.672 to 796.279 mg/L) and TDS difference (753- 3,614 mg/L) that also exceeded the standard in some areas (AlSaadi *et al.*, 2019) Heavy metals were the source of worry in groundwater quality. Mahmoud & Hassan (2021) also reported Pb levels of 0.355–0.509 mg/L above the standards of World Health Organization (WHO). Zinc concentrations ranged from 0.033 to 3.841 mg L–1 and some of the samples exceeded safe levels. Heavy metals indicate industrial and agricultural runoff could be the sources of contamination (Mahmoud & Hassan, 2021).

Recent research has shown that the groundwater in Samarra's water aquifers has vast hydrochemical fluctuations. Al-Tameemi *et al* reported that electrical conductivity varies between 650 and 4,200 µS/cm, reflecting different salinities. The pH ranges from about 7.1 to 8.0. Nevertheless, the values of total dissolved solids (TDS) are rejoicing the limits (water hardness and usability) and may range between 800 and 4,500 mg/L (AL-Tameemi *et al.*,2020) . Ground heavy metal pollution is a great concern of Samara's ground water. Hamza & Younis (2021) reported lead levels varying from 0.25 to 0.65 mg/L, above the WHO drinking water guidelines. Arsenic and cadmium concentrations were similarly detected in trace levels, but exceeded permissible levels in some locations, probably as a result of industrial and agricultural waste (Hamza & Younis, 2021). Biological Quality Micro-organisms contamination is one of the important problems that should be considered in assurance of the groundwater safety. Kareem *et al.* (2019) tested 20 well water from Samarra city for coliform bacteria, which was detected in 40% of samples suggesting for a possible fecal contamination. These results suggest the public health risk and further for effective disinfection prior

than just impacts on humans and aquifers, but also the environment and the economy as a whole. Irrigation with saline groundwater results in the salinization of soils limiting agricultural productivity and may lead to land becoming unfarmable in the long term. The cost related to water purification, the medical treatment of water-related diseases, and rehabilitation of degraded ecosystems represents a significant burden on the scarce resources of the region (Li *et al.*, 2020). Also, the absence of public knowledge about the safe usage and conservation of groundwater makes the situation worse, as several communities unknowingly pollute groundwater by irrational waste disposal and the inept usage of irrigation systems.

Investigations show that the groundwater of Al-Alam district in Salah Al-Din has high values of Total Dissolved Solids (TDS) ranging between 1930 and above 1000 mg/L, which exceed the allowable limit prescribed by WHO and for Iraqi standard. The recorded EC data indicate that values can reach up to 3940 µS/cm which is higher than the normal permissible degree 411 mg/L), both exceeding the WHO maximum permissible limit (250 mg/L), as a result of the dissolution of evaporite minerals like gypsum and halite (Hussain *et al.*, 2022). Groundwater in Al-Alam is also very hard, and has hardness >180 mg/L. Investigation investigated the suitability of ground water for irrigation in Al-Alam using the indices of (EC), (SAR) (Na%), (RSC), (MH), and (PI). Results show that 62% of the samples were found above the acceptable limits of EC for irrigation which SAR and Na% placed some of the ground water under doubt for agricultural purposes (Mahmood & Jassim, 2023). Microbial pollution had been discovered in all tested wells with coliform bacteria density 3-240 cells/100 ml in study on coliform bacteria presence in wells water. The most contaminated samples were from Al-Alam in November which imposed health risks to the consumers (Ali *et al.*, 2022). Some physicochemical studies of groundwater in Balad have been conducted. AlSaadi *et al*1 performed a study to

recharged aquifers (UNESCO, 2021). Longer spells of drought have resulted in low recharge and increase the dissolved constituents concentration, due to lack of dilution and have an adverse impact on water quality (Al-Ansari *et al.*2018) Such climatic conditions, together with over-abstraction, jeopardize the long-term viability of the region's aquifers. Beyond agricultural runoff and domestic wastewater, industrial activities in and around Salah Al-Din are a new relevant source of groundwater pollution. Several oil refineries, chemical plants, and manufacturing facilities are in the governorate and most do not have wastewater treatment infrastructure. Hydrocarbons, heavy metals, and toxic solvents from industrial effluents are commonly discharged straight into the environment or indirectly infiltrate groundwater systems via unlined waste pits and poorly managed disposal sites (Ali & Al-Haidari, 2021). Wastewater from hospitals is disposed to the public network then to central wastewater treatment plants. partly treated effluent is sneak to the groundwater and the some discharged to the sea(Abd Al Satar & Sachit, 2021a). As illustrated, the total of these discharges – and therefore long-term aquifer contamination – is most harmful in regions with high industrial density like Baiji, potentially triggering irreparable ecological damage and public health risks. Salah Al-Din Governorate is geographically situated in the Middle of the Iraqi Tigris River basin, which is the geographical center of Iraq's agricultural and economic activities. Long drought durations, the unevenness of surface water, and the deterioration of river systems as a result of upstream damming and contamination have rendered the wetland areas of Salah Al-Din historically dependent on groundwater resources for its citizens and industries (Ehab Mohammad *et al.*, 2020). This has particularly increased the pressure on aquifer systems that are already overexploited with plenty showing signs of declining quality and quantity. Poor groundwater quality has wider implications

3. Survey of Previous Studies in Salah Al-Din Governorate, Iraq

A variety of natural and human-induced aspects affect the groundwater quality of Salah Al-Din. Sedimentary formations with a more complex hydrogeological structure characterize the geology of the region and influence the mineral composition of groundwater. Constituents, like most iron, manganese, and sulfates, may be released during the dissolution process in the aquifers of rock-water through natural elements (Al-Dabbas *et al.*, 2018). Most importantly, human activities are the source of greater issues. With the unregulated agriculture of chemical fertilizers and pesticides large amount of nitrates and phosphates has increased the levels in groundwaters. Likewise, the release of untreated sewage and solid waste, together with leaching from septic tanks, have contributed to microbial pollutants and heavy metals in to ecosystems (Mahdi & Al-Jaberi, 2021). The differences in groundwater quality observed in Salah Al-Din are related to numerous hydrogeological and land use factors. Shallow aquifers which are at high risk of pollution due to their proximity to agricultural land and populated areas, are concentrated particularly in southern and central areas of the governorate. Shallower aquifers, however, provide more protection, but are also far less expensive to access and monitor (Ismail *et al.*, 2023). In addition, the absence of advanced wastewater treatment infrastructure in the majority of Salah Al-Din has resulted in the release of pollutants directly or indirectly into the subsurface, an action which threatens the safety of groundwater intended for drinking purposes. Beyond anthropogenic pressures, climate change has increased the risk to groundwater in Salah Al-Din. The last 20 years have shown more intensively the rise of temperature as well as the fall of yearly precipitation in Iraq, either of which has led to the decreasing of renewable water resources and to a larger dependency on non-renewable or slowly

RSC measures the ratio of carbonate and bicarbonate ions versus calcium and magnesium ions. High levels of carbonate and bicarbonate in comparison to calcium and magnesium, could lead to the precipitation of calcium and magnesium. This negatively affect the combined relative sodium, soil dispersion, infiltration, and soil structure (Richards, 1954). High RSC waters will lead to the addition of sodium in soil, which has negative influences on soil permeability and crop yield. The problem is major in dryland areas where leaching effect is less and irrigation is an integral part of agriculture.

2.11 Water Quality Index (WQI) of Groundwater for Agriculture

Water Quality Index (WQI) for agricultural use is an important indicator for assessing the suitability of the ground water for irrigation. The index is based on the analysis of selected physio-chemical properties which influence soil structure and crop health as well as long term agricultural productivity. It usually involved measurements of EC, SAR, RSC, TDS, chlorides, nitrates, and boron.

High EC and TDS of water would cause soil salinity problems, while high SAR and RSC would destabilize soil structure and reduce its permeability, resulting in unsuitable water for crop growing. Thus, evaluation of the WQI is significant to the farmers and planners for safer irrigation water and also to make rectifications, if need be.

WQI is a single value summary of complex water quality data in water resource management for agriculture, which is easy to interpret and suitable for decision making. This is a significant issue in sustainable agriculture, especially in places where there is contamination of groundwater or over extraction (Ramesh &Elango, 2012).

environment, whilst magnesium can also contribute to the structure and fertility of a soil (Lindsay, 1979).

2.9 Sodium Adsorption Ratio (SAR)

The Sodium Adsorption Ratio (SAR) is used to assess the relative concentration of sodium (Na^+) to calcium (Ca^{2+}) and magnesium (Mg^{2+}) in irrigation water. It is a crucial indicator of the potential for sodium to negatively impact soil structure (Ayers & Westcot, 1985; USDA-NRCS, 1997).

Formula:

$$SAR = Na / \sqrt{(Ca + Mg)/2}$$

Where all ions are measured in milliequivalents per liter (meq/L).

Importance of SAR in Irrigation

In water of high SAR, displacement of calcium and magnesium ions can occur from the surface of soil particles resulting in soil dispersion, crusting, lower permeability and poor infiltration (Ayers & Westcot, 1985). This can severely

such as affect plant roots growth and water availability.

2.10 Residual Sodium Carbonate (RSC) in Irrigation Water

RSC value is an important criterion for determining the suitability of water for irrigation, specifically its effect on soil structure and percolation properties. It is the sequence of value calculated as:

$$RSC=(CO_3^{2-}+HCO^{3-}) - (Ca_2^{+}+Mg_2^{+})$$

Where all ionic concentrations are expressed in milliequivalents per liter (meq/L).

Calcium (Ca)

It is also indispensable in the maintenance of cell wall structure and integrity in plants. It participates in the production of new cells and in root growth. Sufficient calcium in ground water can improve the soil structure thereby better drainage and aeration, both of which are crucial for optimal root health (Marschner, 2012). Calcium also contributes to reduce the impact of soil salinity that can negatively

Magnesium (Mg)

A central part of chlorophyll, the molecule that absorbs sunlight and creates energy in green leaves, relies on magnesium. It is involved in energy transfer and enzyme activation in plants. Mycorrhiza induced magnesium uptake from groundwater could enhance photosynthetic efficiency with resultant increase in crop yield (Marschner, 2012). Furthermore, the uptake of other important nutrients (e.g., N and P) is facilitated by magnesium and thereby has a potential to improve plant yields and quality (Cakmak, 2008).

Effects on Agriculture

The levels of calcium and magnesium in the groundwater may play a critical role in soil fertility and fruit crops. Soils containing these elements in the correct balance usually result in better plant health, more disease resistance and better nutrient uptake. In contrast to optimal availability, either one of these elements may be insufficient and resulting in physiologic disorders in plants, as blossom end rot in tomato because of diminishing calcium and interveinal chlorosis in crops such as corn and bean due to decreasing magnesium (White & Broadley, 2003). In addition, the availability of Ca and Mg affects soil pH and CEC and therefore nutrient availability. Calcium can assist in the neutralization of an acid soil

Sulfate is an essential source of sulfur, a macronutrient required for protein synthesis, enzyme activity, and chlorophyll formation in plants (Hawkesford, 2000). When present in moderate concentrations in irrigation water, sulfate contributes positively to crop growth and productivity, especially in sulfur-deficient soils.

However, elevated sulfate levels in irrigation water may lead to several challenges. High sulfate concentrations can contribute to the formation of saline or sodic soils, particularly in arid and semi-arid regions. Sulfate salts, such as sodium sulfate, can increase the total salinity of the soil solution, reducing water availability to plants through osmotic stress (Ayers & Westcot, 1985). Additionally, excessive sulfate can lead to nutrient imbalances by interacting with calcium and magnesium, potentially causing deficiencies or toxicities depending on soil and crop conditions.

In livestock production, high levels of sulfate in drinking water (>500 mg/L) may have negative effects, particularly in young animals, with symptoms such as decreased feed consumption, diarrhea or, in extreme cases, sulfur poisoning, which may cause polio encephalomalacia in ruminants (Loneragan *et al.*, 2001). Appropriate water quality evaluation and management decisions such as blending water sources or amendments to the soil are crucial to minimize the potential negative impacts due to high sulfate concentrations in agricultural environments.

2.8 Calcium (Ca) and Magnesium (Mg) in water

Calcium (Ca) and magnesium (Mg) are major nutrients of groundwater with a crucial role in agricultural production. Both are essential elements affecting many physiological and biochemical processes in plants.

soil which can cause deficiency and unbalanced nutrients and soil structure (Ayers & Westcot, 1985). The resulting decrease in the relative levels of calcium and magnesium increases the concentration of sodium (Na^+), leading to soil sodicity, a property that leads to deteriorating soil structure, decreased soil permeability, and restricted root development. An important index to assess such a risk is RSC (Residual Sodium Carbonate), which is determined from carbonate, bicarbonate, calcium and magnesium concentrations. High RSC values indicate a greater risk of soil degradation when such water is used for irrigation (Rhoades, Kandiah, & Mashali, 1992). a derived indicator from HCO_3^- and CO_3^{2-} levels — as an important hazard factor:

$$\mathbf{RSC=(CO_3^{2-}+HCO_3^-) - (Ca^{2+}+Mg^{2+})}$$

RSC > 2.5 me/L is considered a severe hazard.

Bicarbonates can also affect plant nutrient uptake by altering soil pH. Elevated soil pH caused by bicarbonate accumulation may reduce the availability of micronutrients such as iron, zinc, and manganese, which are critical for plant metabolism and development (Gupta & Gupta, 1987).

To mitigate these effects, farmers may use soil amendments such as gypsum (calcium sulfate) to restore calcium levels, improve soil structure, and reduce sodium hazards.

2.7 Sulfate (SO_4^{2-}) in Water

Sulfate (SO_4^{2-}) is a naturally occurring anion in soil and water, derived from the weathering of sulfate-containing minerals such as gypsum and from anthropogenic sources like industrial discharge and fertilizer runoff. In agricultural water, sulfate plays a dual role—acting as a necessary plant nutrient while also potentially posing risks when present in excess.

into surface and ground water is usually due to overuse of fertilizers or inefficient irrigation. This pollution causes a decrease in the quality of water, rendering it unfit for human consumption, and can also contribute to eutrophication in adjacent aquatic systems (Rivett *et al.*, 2008). Eutrophication is the process of nutrient loading, which leads to excessive algal production, causing reduction of oxygen and damage to the organisms living in the water. On one hand, contaminated water with high nitrate content can be welcome for agriculture. Low concentrations nitrate in the irrigation water, on the other hand, can act as an additional source of nitrogen for crop uptake which may reduce fertilization requirements. However, high levels of nitrate in irrigation water may cause nutrient imbalance, soil salinity and nitrate toxicity, which is especially dangerous in forage crops utilized as fodder by livestock (Addiscott, 2005).

The control of nitrate in agricultural water systems is contingent on integrated strategies including the practice of nutrient budgeting, adoption of cover crops, and improved irrigation efficiency, and establishment of buffer areas to minimize runoff. Such strategies could help farmers remain productive even as they reduce the environmental toll from nitrate pollution.

2.6 Carbonate (CO_3^{2-}) and Bicarbonate (HCO_3^-) in Water

Carbonate (CO_3^{2-}) and bicarbonate (HCO_3^-) ions present in irrigation water are the most common pH determining factors in irrigation water. These ions derive from the dissolution of mineral deposits (like limestone), and may have a great impact on soil chemistry and agriculture productivity.

Excessive 'hard conditions' in field irrigation water can result in calcium (Ca^{2+}) and magnesium (Mg^{2+}) being precipitated—essential nutrients for plant health. This also diminish the availability of these nutrients in the

to maintain soil productivity in the long run. However, adequate practices which attenuate the adverse effects of salinity include leaching practices, soil amendments (such as gypsum for sodium) and appropriate salt-tolerant crop choices (Grattan & Grieve, 1999).

Sodium (Na^+): One element of the first group that is even more problematic for agriculture is sodium (Na^+) due to its influence on soil structure. Excess sodium can disperse soil particles, impeding infiltration & permeability, causing problems with root development and waterlogging. Sodium also displaces important plant nutrients such as potassium and calcium. The impact of sodium (Na) is often evaluated using the Sodium Adsorption Ratio (SAR) for water quality assessments (Grattan & Grieve, 1999).

Potassium (K^+): It's an essential macronutrient in plant metabolism, playing a key role in enzyme activation, regulation of water and photosynthesis. However, at high levels it can interfere with uptake of magnesium and calcium, causing nutrient deficiencies and poor crops, it is generally believed to be less toxic than sodium or chloride. Irrigation water potassium levels are generally not harmful by themselves unless they are enhanced by fertilizer leaching or, industrial sources (Grattan & Grieve, 1999)

2.5 Nitrate (NO_3^-) in Water

Nitrate (NO_3^-) is a common form of nitrogen found in agricultural runoff and groundwater due to the widespread use of nitrogen-based fertilizers. While nitrate is essential for plant growth, its presence in water—particularly in excess—can pose significant challenges to both environmental quality and agricultural sustainability. In agricultural areas, nitrate seeping

stunted growth. Acidic water may also damage the irrigation equipment, raising the costs of maintenance and limiting the lifespan of the systems (Horneck *et al.*, 2007). In contrast, alkaline water, i.e. water with pH above 8.4, can reduce the availability of nutrient such as phosphorus, iron, and manganese. Such deficiencies typically manifest as chlorosis, stunted growth and reduced yields. Alkaline water often has a high concentration of bicarbonate and carbonate ions, which will exacerbate soil structural problems and sodium problems (Horneck *et al.*, 2007) To avoid the above problems, farmers must regularly measure and control the pH of the irrigation water, using acidifying or neutralizing agents when required. This preserve soil fertility and ensures that the assets of the soil properly use, leading to supporting sustainable crop production (Horneck *et al.*, 2007).

2.4 Chloride (Cl^-), Sodium (Na^+), and Potassium (K^+) Ions in Water

Chloride (Cl^-), sodium (Na^+), and potassium (K^+), are often present in the irrigation water and are the major ions containing salts that are responsible for affecting the growth of plants and the health of the soil. Although these ions are necessary for plant growth in small quantities, over the threshold of tolerance they become harmful to plants or can persist in the soil for long periods (Grattan & Grieve, 1999).

Chloride (Cl^-): It's a required micronutrient; but at high concentration, it induces marginal leaf burn and abscission in the leaves as well as yield loss in crops, especially in chloride susceptible species like avocado, grape and citrus. Under high evapotranspiration conditions, the impacts are amplified as chloride is more readily stored in plant tissues. The toxicity due to the excess of these ions in irrigation water can still be monitored and managed

which means lesser crop growth leading to lower yields. In addition, some ions that comprise TDS, including sodium and chloride, may be harmful to sensitive crops when their concentrations exceed threshold levels. Prolonged salt exposure may change the structure and permeability of soils, especially those with high sodium content, resulting in decreased infiltration and aeration. High TDS levels can also affect the application of fertilizers and pesticides, as water with dissolved solids above a certain level can also affect the solubility or cause adverse reactions between the applied chemical and the water at higher levels. High TDS in irrigation can also damage soil, so, monitoring TDS in irrigation is highly critical to maintain soil health as well as to achieve maximum crop production. The Food and Agriculture Organization (FAO) indicate that irrigation water contains TDS 2,000 mg/L is unsuited to crops and must be managed very carefully (For more arranging but still far reader) (Ayers & Westcot, 1985).

2.3 pH of Water

The pH of irrigation water is an important agricultural characteristic determining the soil chemistry and availability of nutrients vital for plant growth. The pH of water that is outside the most productive band can affect the health and yield of the crop. According to Horneck *et al*. (2007), the pH in the saturated area of the soil should fall within the range of 6.5-8.4, as extreme acidity or alkalinity would negatively influence nutrient uptake and structure of the soil, therefore deviating from the optimum soil pH for most agricultural crops. Water for irrigation with acid (pH < 6.5) can cause soil acidification, with documented higher solubility of toxic elements as Al and Mn. These substances can be toxic to plants, causing root destruction and

Salinity tolerance is different for various crops. There is a wide range, with broadly salt-tolerant crops (like barley and cotton) on one end of the continuum, and highly salt-sensitive crops (like beans and strawberries) on the other. EC monitoring enables crop selection, leaching application and irrigation scheduling to avoid salt build-up in the root zone. In addition, the use of high-EC water for extended periods may also cause soil degradation, especially in heavy-textured poorly-drained soils (where salts may precipitate and accumulate over time). This might require further management tools, including soil amendments or drainage systems, to keep the soil productive. Accordingly, EC is an important parameter to test the quality of irrigation water and assists farmers for sustainable agricultural implementation and soil salinity prevention.

2.2 Total Dissolved Solids (TDS)

Total Dissolved Solids (TDS) are, by definition, the total concentration of all dissolved substances (both organic and inorganic) in water, including minerals, salts, and metals. They are found in the ionic form, which includes calcium, magnesium, sodium, potassium, chlorides, sulfates and bicarbonates. Total Dissolved Solids (TDS): Expressed in milligrams per liter (mg/L) and one of the most key measures of water quality for agricultural purposes.

That TDS can cause salinity problems with irrigation water in agriculture. If high TDS water is continuously used, salts will build up in the soil, particularly in saline-prone areas and dry climates where drainage is inadequate. It decreases osmotic potential of the soil, making it more difficult for roots to absorb water, even when moisture appears adequate. This leads to water stress (Ayers & Westcot, 1985) as well as crop stress

2. Water quality parameters and their effects on agriculture

Various systematic approaches of measurement, sampling and laboratory analysis is available to assess the groundwater quality for irrigation. The choice of water quality indices is only the first step, as IWQI is constructed around different physicochemical parameters for the evaluation of suitability of water for irrigation. Factors such as EC, sodium adsorption ratio (SAR) and Na% are constraining. (Ajmal *et al.*, 2023). Knowledge of the physical, chemical and biological parameters of water is very important before using it for various purposes, which inverts or describes water quality(Abd Al Satar & Sachit, 2021)

2.1 Electrical conductivity (EC)

EC is one of the most important indices of salinity in water and soil, which reflects the contents of the dissolved salts, including sodium, calcium, magnesium, chloride, sulfate, and bicarbonate. In agriculture, EC is commonly used to evaluate the suitability of irrigation water and salinity status of soils and are the most essential factors for improvement of plant and crop production.

Irrigation with saline water is treated as that water has sure effects on plants. So high EC value in irrigation water belongs to high salt contents. High-EC water can lead to elevated soil salinities, which can impose osmotic stress in plants and limit water extraction even when soil moisture is seemingly adequate. As a result, it can reduce germination rates, stunt growth, and freeze yields (Ayers & Westcot, 1985).

1.Introduction

Groundwater quality is an integral part of agricultural productivity as it has a direct impact on the soil and subsequently on crop production. Polluted groundwater would adversely affect crops and ecosystems, which creates the need to control water quality. (Saeed, 2014). Groundwater quality parameters are defined in both physical, chemical, and biological terms. Temperature, turbidity, and electrical conductivity are the physical parameters that may affect the usability of the water. Chemical parameters include nutrients, salts, and pollutants; biological parameters include microorganisms that can affect the safety and quality of water. (Acharya *et al.*, 2010). Sustainable and resilient agricultural practices depend on freshwater resources, the management of which will be vital in the context of global change. These include measures to promote water efficient use, risk reduction of contamination and recharge of groundwater enhancement methods. Decision-makers on the policy front, among food producers and across the research spectrum should be working to seek out modern technology and best practices that work in equilibrium with sustainable agriculture, and environmental stewardship." (Varela-Ortega *et al.*, 2011). In short, groundwater is essential for agriculture, especially in areas where surface water is scarce. They are required for the continuity of these resources and the sustainability of agricultural systems (Srivastav *et al.*, 2021).

Hence; this review focuses on the effect of groundwater quality parameters on plant and crop yield when being irrigated by groundwater, it emphasizes the recent studies that investigated this in the past.

المستخلص

تركز هذه الدراسة على تأثير الخصائص الفيزيائية والكيميائية للمياه الجوفية على النباتات والمحاصيل الزراعية، وذلك بعد مراجعة الدراسات السابقة المتعلقة بهذا الموضوع. وتؤكد النتائج على ضرورة معالجة المياه الجوفية قبل استخدامها في الري لتحقيق زراعة مستدامة. كما يُمكن تحديد نوع المحاصيل التي يمكن زراعتها في الأراضي المروية بمياه الابار، وفقًا لخصائص مياه البئر والغلة النباتية لكل محصول. وتقدم الدراسة أيضًا بعض الأفكار حول معايير جودة مياه الري مثل الموصلية الكهربائية (EC)، والمواد الصلبة الذائبة الكلية (TDS)، والرقم الهيدروجيني، وأيونات الكلوريد (Cl^-)، والصوديوم (Na^+)، والبوتاسيوم (K^+)، وأيونات النترات (NO_3^-)، والكربونات (CO_3^{2-})، والبيكربونات (HCO_3^-)، والكبريتات (SO_4^{2-})، والكالسيوم (Ca) والمغنيسيوم (Mg)، ونسبة امتزاز الصوديوم (SAR)، وكربونات الصوديوم المتبقية (RSC)، ومؤشر جودة المياه (WQI) وتأثيرها على المحاصيل والتربة. كما تم تقديم العديد من الدراسات حول جودة المياه الجوفية في محافظة صلاح الدين وملاءمتها للري. كما يُؤكد البحث على أهمية حماية المياه الجوفية من التلوث الناتج عن التركيزات المفرطة من العناصر الضارة، والتي تُؤثر سلبًا على النباتات والتربة، وبالتالي على الأمن الغذائي. باختصار، تم استنتاج ان للمياه الجوفية أهمية كبيرة في تطبيقات مُختلفة ومنها الزراعة، وتتطلب إدارة مُستدامة. لذلك، يجب اتباع استراتيجيات مُحكمة ومخططة جيدًا قبل إنشاء أي مشروع بالقرب من مصادر المياه الجوفية، وذلك للحفاظ عليها وحماية البيئة المُحيطة بها، بالإضافة الى ضرورة معالجتها قبل استخدامها للري.

الكلمات المفتاحية: معاملات جودة المياه الجوفية; الموصلية الكهربائية; غلة المحصول; الزراعة; محافظة صلاح الدين

Abstract

The focus of this study is the effect of the physical and chemical characteristics of groundwaters on plants and agricultural crops, following the last studies on this matter. The results underscore the necessity for groundwater treatment before its utilization in irrigation for sustainable agricultural farming. It is also possible to identify which crop type can be grown in well-watered land, according to the characteristics of irrigation water and the yield of each crop. The study also introduces some ideas about irrigation water quality parameters such as Electrical conductivity (EC), Total Dissolved Solids (TDS), pH, Chloride (Cl^-), Sodium (Na^+) and Potassium (K^+) Ions, Nitrate (NO_3^-), Carbonate (CO_3^{2-}) and Bicarbonate (HCO_3^-) Ions, Sulfate (SO_4^{2-}), Calcium (Ca) and Magnesium (Mg), Sodium Adsorption Ratio (SAR), Residual Sodium Carbonate (RSC) and Water Quality Index (WQI) and their effect on crops and soil. Many of the series of studies on the groundwater quality in Salah Al-Din governorate and its suitability for irrigation were presented in this study. The research also reinforces the importance of the protection of groundwater against pollution by excessive concentrations of harmful elements with adverse effects on plants, soil, and, ultimately, food security. In summary, found that the groundwater is of considerable importance in various applications specially in agriculture, and requires sustainable management. Therefore, thorough and well-planned strategies must be adopted before establishing any project near groundwater sources, in order to preserve them and protect the surrounding environment, and it's necessary to be treated before using in irrigation.

key words: groundwater quality parameters; EC; crop yield; agriculture; Salah Al-Din Governorate

Groundwater Quality Analyses for Irrigation Purposes in Salah Al-Din, Iraq: A Review/ Review article

Noor A. Radhi [1*], Dawood E. Sachit [2] and Abdul–Sahib T. Al–Madhhachi [3]

1 Researcher, Environmental Engineering, Department of Environmental Engineering, College of Engineering, Mustansiriyah University, Baghdad, Iraq.

2 Assist. Prof. Dr., Department of Environmental Engineering, College of Engineering, Mustansiriyah University, Baghdad, Iraq.

3 Prof. Dr., Department of Water Resources Engineering, College of Engineering, Mustansiriyah University, Baghdad, Iraq.

* Corresponding Author: noorradhi90@uomustansiriyah.edu.iq

تحليل جودة المياه الجوفية لأغراض الري في محافظة صلاح الدين، العراق: دراسة مرجعية

نور عبد الرزاق راضي[1*]، داوود عيسى ساجت [2]، عبد الصاحب توفيق المذحجي[3]

1. بكالوريوس علوم في هندسة البيئة، قسم هندسة البيئة، كلية الهندسة، الجامعة المستنصرية، بغداد، العراق
2. أ.م. د.، قسم هندسة البيئة، كلية الهندسة، الجامعة المستنصرية، بغداد، العراق
3. أ. د.، قسم هندسة الموارد المائية، كلية الهندسة، الجامعة المستنصرية، بغداد، العراق

* المؤلف المراسل: noorradhi90@uomustansiriyah.edu.iq

[20] Noshad, Z., Javaid, N., Saba, T., Wadud, Z., Saleem, M. Q., Alzahrani, M. E., & Sheta, O. E. (2019). Fault detection in wireless sensor networks through the random forest classifier. Sensors, 19(7), 1568.

[21] Moorthy, U., & Gandhi, U. D. (2021). A novel optimal feature selection technique for medical data classification using ANOVA based whale optimization. Journal of Ambient Intelligence and Humanized Computing, 12(3), 3527-3538.

[22] Yin, Y., Jang-Jaccard, J., Sabrina, F., & Kwak, J. (2022). Improving Multilayer-Perceptron (MLP)-based Network Anomaly Detection with Birch Clustering on CICIDS-2017 Dataset. arXiv preprint arXiv:2208.09711.

[23] Papamartzivanos, D., Mármol, F. G., & Kambourakis, G. (2019). Introducing deep learning self-adaptive misuse network intrusion detection systems. IEEE Access, 7, 13546-13560.

[24] Sahoo, K., Samal, A. K., Pramanik, J., & Pani, S. K. (2019). Exploratory data analysis using Python. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(12), 2019.

[25] Hao, W. T., Lu, Y., Dong, R. H., Shui, Y. L., & Zhang, Q. Y. (2022). Adaptive Intrusion Detection Model Based on CNN and C5.0 Classifier. International Journal of Network Security, 24(4), 648-660.

[26] Shukla, A. K. (2021). Detection of anomaly intrusion utilizing self-adaptive grasshopper optimization algorithm. Neural Computing and Applications, 33(13), 7541-7561.

[27] Bhavsar, M., Roy, K., Liu, Z., Kelly, J., & Gokaraju, B. (2022). Intrusion-Based Attack Detection Using Machine Learning Techniques for Connected Autonomous Vehicle. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (pp. 505-515). Springer, Cham.

[28] AKGUN, Devrim; HIZAL, Selman; CAVUSOGLU, Unal. A new DDoS attacks intrusion detection model based on deep learning for cybersecurity. Computers & Security, 2022, 118: 102748.

[29] SAMBANGI, Swathi; GONDI, Lakshmeeswari. A machine learning approach for DDoS (distributed denial of service) attack detection using multiple linear regression. In: Proceedings. MDPI, 2020. p. 51.

[30] SAHOO, Kshira Sagar, *et al.* A machine learning approach for predicting DDoS traffic in software defined networks. In: 2018 International Conference on Information Technology (ICIT). IEEE, 2018. p. 199-203.

[31] MUSTAPHA, Ali, *et al.* Detecting DDoS attacks using adversarial neural network. Computers & Security, 2023, 127: 103117.

[8] Oqaily, A., Jarraya, Y., Wang, L., Pourzandi, M., & Majumdar, S. (2022). MLFM: Machine Learning Meets Formal Method for Faster Identification of Security Breaches in Network Functions Virtualization (NFV). In European Symposium on Research in Computer Security (pp. 466-489). Springer, Cham.

[9] Saeed, M. M. (2022). A real-time adaptive network intrusion detection for streaming data: a hybrid approach. Neural Computing and Applications, 34(8), 6227-6240.

[10] Latif, S., Zou, Z., Idrees, Z., & Ahmad, J. (2020). A novel attack detection scheme for the industrial internet of things using a lightweight random neural network. IEEE Access, 8, 89337-89350.

[11] Chou, D., & Jiang, M. (2020). Data-driven network intrusion detection: A taxonomy of challenges and methods. arXiv preprint arXiv:2009.07352.

[12] Alnaim, A., Alwakeel, A., & Fernandez, E. B. (2019, August). A Misuse Pattern for Compromising VMs via Virtual Machine Escape in NFV. In Proceedings of the 14th International Conference on Availability, Reliability and Security (pp. 1-6).

[13] Alwakeel, A. M., Alnaim, A. K., & Fernandez, E. B. (2019, May). Toward a Reference Architecture for NFV. In 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS) (pp. 1-6). IEEE.

[14] Fagroud, F. Z., Ajallouda, L., Toumi, H., Achtaich, K., & El Filali, S. (2020). IOT search engines: exploratory data analysis. Procedia Computer Science, 175, 572-577.

[15] Patel, H., Guttula, S., Mittal, R. S., Manwani, N., Berti-Equille, L., & Manatkar, A. (2022, August). Advances in exploratory data analysis, visualisation and quality for data centric AI systems. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 4814-4815).

[16] Almomani, O. (2020). A feature selection model for network intrusion detection system based on PSO, GWO, FFA and GA algorithms. Symmetry, 12(6), 1046.

[17] Shastri, Sourabh, and Vibhakar Mansotra. "Kdd-based decision making: a conceptual framework model for maternal health and child immunization databases." Advances in computer communication and computational sciences. Springer, Singapore, 2019. 243-253.

[18] Belgrana, F. Z., Benamrane, N., Hamaida, M. A., Chaabani, A. M., & Taleb-Ahmed, A. (2021, January). Network intrusion detection system using neural network and condensed nearest neighbors with selection of NSL-KDD influencing features. In 2020 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS) (pp. 23-29). IEEE.

[19] Dashpurev, B., Bendix, J., & Lehnert, L. W. (2020). Monitoring oil exploitation infrastructure and dirt roads with object-based image analysis and random forest in the Eastern Mongolian Steppe. Remote Sensing, 12(1), 144.

whether they are two-class or multi-class, and it can adapt to changes in the network environment.

6) The proposed system's results are compared with related work, and the best accuracy is recorded with our proposed system.
7) The KDD dataset was used to test and evaluate the proposed system based on performance metrics of accuracy.
8) The proposed model is capable of defending against the most common and dangerous attack, which could be NFV, which is a misuse attack.

References

[1] R. Diesch, M. Pfaff, & H. Krcmar" A comprehensive model of information security factors for decision-makers" Computers & Security, Vol. 92, 2020.

[2] A. Tchernykh, U. Schwiegelsohn,E.G. Talbi & M. Babenko, "Towards understanding uncertainty in cloud computing with risks of confidentiality, integrity, and availability" Journal of Computational Science, Volume 36, September 2019, 100581.

[3] A. A. Ahmed & N. B. Al Dabbagh," Web Attacks and Defenses", Journal of Education & Science, Vol, 32, No: 2, 2023 (114-127.

[4] N. Alhebaishi, L. Wang & S. Jajodia," Modeling and Mitigating Security Threats in Network Functions Virtualization (NFV)", IFIP Annual Conference on Data and Applications Security and Privacy (pp. 3-23). Springer, June , 2020.

[5] F. Paganelli, P. Cappanera & G. Cuffaro, " Tenant-defined service function chaining in a multi-site network slice", Future Generation Computer Systems, Volume 121, August 2021, Pages 1-18.

[6] S. Cherrared, I. Sofiane, F. Eric, G. Gregor, & G. B. Y. Imen, "A survey of fault management in network virtualization environments: Challenges and solutions." IEEE Transactions on Network and Service Management 16, no. 4 (2019): 1537-1551.

[7] I. Alam, K. Sharif, F.Li , Z. Latif, M. M. Karim, N.Nour, S. Biswas, & Y. Wang, "IoT virtualization: A survey of software definition & function virtualization techniques for internet of things." arXiv preprint arXiv, 2019, 1902.10910.

6. Conclusions

The suggested misuse attack detection model improves the ability of the misuse attack detection model in a Network Function Virtualization network to be able to adapt to the dynamic changes in the network environment. It is based on an NFV and a Random Forest, Niave Bayse classifier. Many algorithms are covered in this work to secure and protect the NFV against the most common and dangerous attacks that could target it.

The proposed system is applicable in NFV architecture, Cloud Computing, IoT environment, Web server and employed to discloser the Misuse attack traffics features that mislaid among the other traffics and features. The main focus of this work has been as follows:

1) This work proposed an early and accurate system that has the ability to defense against Misuse attack targeting.
2) The great challenge of this work is to distinguish between Abuses trafficking, which mislead and normal trafficking.
3) Random Forest and Niave Bayse were used to base the proposed system on the most effective machine learning algorithms.
4) The dataset is the most important part of testing and evaluating the performance of the proposed system. The proposed system's performance was tested and evaluated using the KDD dataset in this work.
5) According to experimental findings, the machine-learning algorithms has a strong detection performance in terms of detection accuracy, since 99.9% accuracy is achieved for Random Forest and since 99.5% for Niave Bayse .The suggested method's detection accuracy is higher than the most recent techniques,

been tested and assessed using the KDD dataset. The dataset is divided into two segments, 30% of the data is used for testing, while 70% is used for training.

According to the data in Table 4, the proposed method detects more misuse attacks in the KDD dataset than other methods in multi-classification task detection. The suggested method's detection accuracy is higher than the most recent techniques; it is capable of adapting to changes in the network environment, Regardless of whether it is a two-class or multi-class.

Table 4. Compare our Proposed Model with previous Work.

Model	Techniques	Accuracy%
Goel *et al.* [22]	Classification-Based Association (CBA)	97.4
Papamartzivanos *et al.* [23]	Self-adaptive and autonomous misuse of IDS	84
Wen-Tao Hao *et al.*[25]	convolutional neural network (CNN) and C5.0 classifier	98
Shukla *et al.* [26]	opposition self-adaptive grasshopper optimization	99.35
Bhavsar *et al.*[27]	a system for detecting intrusions	98
Akgn *et al.*[28]	The Long Short-Term Memory (LSTM) model	99.2
Sambagi *et al.*[29]	Utilized information gain and regression analysis to learn the ensemble of feature selection	97.86
Sahoo *et al.*[30]	Linear regression, KNN, Nave Bayes, Decision Tree, Random Forest, SVM, and ANN are all available	98.64
Mustapha *et al.*[31]	The IDS can be improved by adding another LSTM model that is responsible for blocking adversarial samples	91.75
our proposed model	Random Forest Niave Bayse	99.98 99.5

5. Analysis and discussion

The control and prevention of incoming abnormal traffic are too difficult a task that could target information security developers. Normally, an accurate and rapid detection system for controlling the attack traffic is required. Even after many methods were used and developed to identify and control the attack traffic. Most of these are not detectable with the best accuracy. Furthermore, NFV may be the victim of several types of attacks, and flood attacks are the most frequent and the most dangerous. Flooding attacks are divided into two main types: misuse and DDoS. Both of their attempts to overwhelm the NFV with traffic and consume system resources failed. This work, emphasizes misuse attack detection in the early stages.

The proposed model used the most common and effective machine learning techniques, which is random forest and Niave Bayse. The proposed system has been tested and evaluated by using the KDD dataset.

In the literature, the majority of methods indicate categorization accuracy ranges between 60% and 92%. Those with better accuracy were typically tested using a particular attack traffic type. When other traffic kinds are introduced or other sorts of attacks, such as DDoS, are taken into consideration, those solutions cannot continue to function as well.

Additionally, the proposed model based on uses the KDD benchmarking dataset to discriminate, extract, and identify the elements of the misuse attack, as well as to distinguish its traffic from other forms of traffic. The design, implementation, testing, and evaluation of effective machine learning approaches that been used for intrusion detection systems make up the second part. These formulas belong to the Random Forest family and Niave Bayse. Furthermore, the machine-learning algorithm's performance has

4.4 Results of Classification

The training stage's misused datasets contributed to 70% of the data in the classification stage, and the testing stage was where 30% came from. The proposed system uses the Random Fours classification algorithm, and the results are based on the accuracy ratio, so this section illustrates the results of the random forest classification (4.4.1) and the analysis and discussion in the subsections (4.4.2).

4.4.1 Results of Random Forest Classification

This section will show the results of the random forest classification algorithm as illustrated in algorithm 2, based on values of accuracy rate using equation (11), with all cases provided in the proposed system as shown in figure 4.

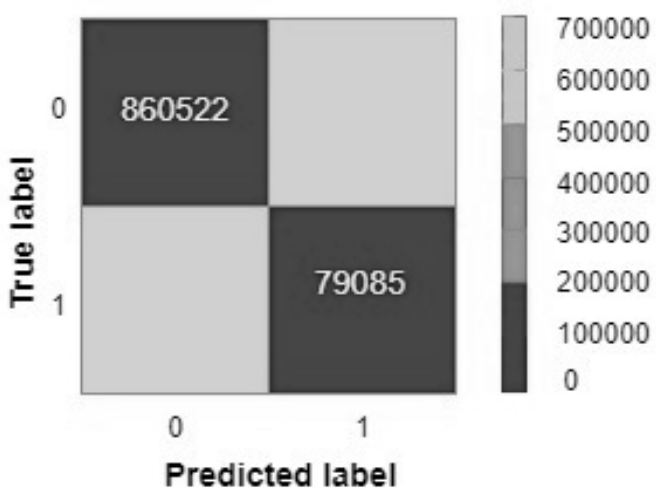


Figure 4. Results of accuracy rate for Random forest classification.

Table 2. First step of one hot encoding algorithm.

#	Feature	Number of categories	
		training	testing
0	protocol type	3	3
1	Service	70	65
2	Flag	11	10
3	label'	23	2

However, quite a range of factors has been taken into consideration, such as Data Quality and Quantity, Data Availability, and Gaps in the Data.

4.3 Feature Selection using PSO-Test

After using the feature one hot encoding and normalization technique, the number of features is Train: Dimensions of Misuse (1431357, 117) and Test: Dimensions of Misuse (1431357, 117). Table 3 displays the results of selecting features.

Table 3. PSO F-test results for feature selection.

[logged_in,
Count,
Serror_rate,
Stv_serror_rate,
Same_ Stv_rate,
Dst_host_stv_Count,
Dst_host_ Same_ Stv_rate,
Dst_host_ Serror_rate,
Dst_host_ Stv_serror_rate,
Service_private,
Flag_S0,
Flag_SF]

4.The Results

This section, results from the proposed system that has been designed to defend against misuse attacks targeting NFV are presented and discussed. The outcomes of each stage of the proposed model are shown below:

4.1 Import Dataset

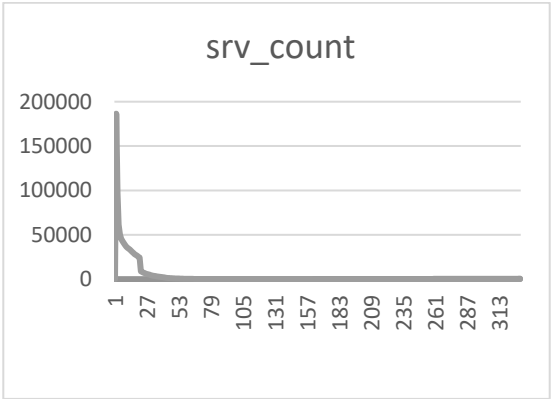
The first step proposes a system load data set and divides the dataset into parts (70% training and 30% testing), where the dimensions of the training set are (211979, 50) and the dimensions of the test set are (325498, 33).

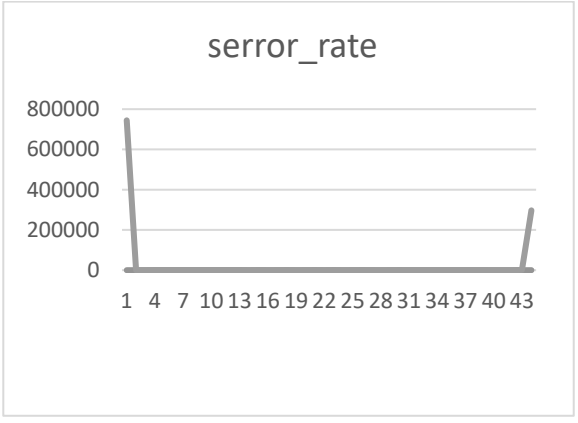
4.2 Data Preprocessing using One Hot Encoding Technique

Furthermore, according to the previous chapters' discussions, Hot Encoding is employed to convert all features into binary. Here is the One-Hot encoding procedure:

- Input [matrix of integers]
- Processing [Denoting values of features]
- Output [Sparse matrix, one feature]

In case of we assume the input features take values in range (0, n_ values). The first step in one hot encoding technique is to identify categorical features, as illustrated in Table 2. There are only four features that have symbols or text, which are protocol type, flag, service, and label.





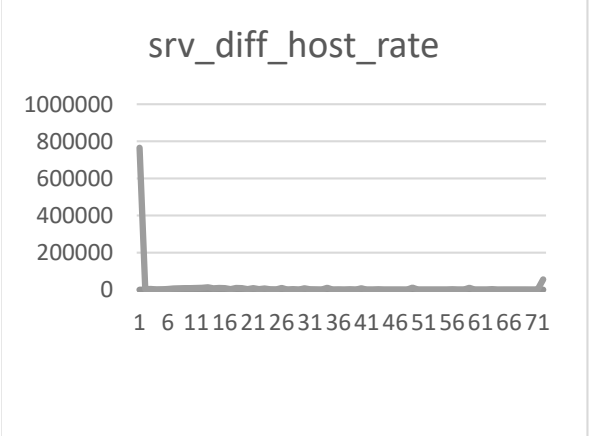
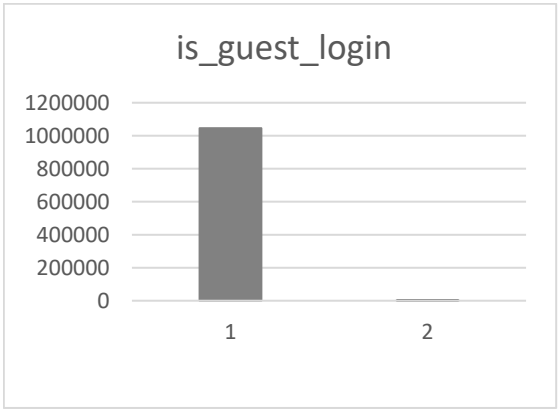
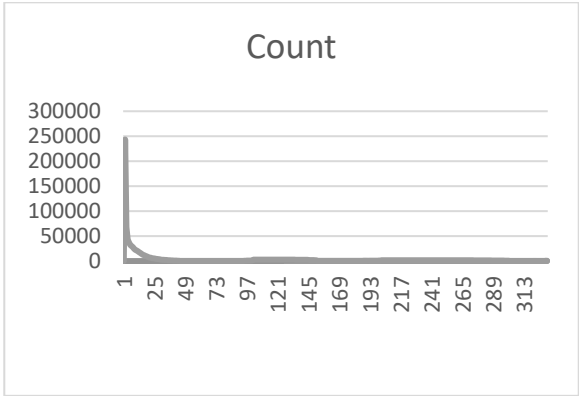
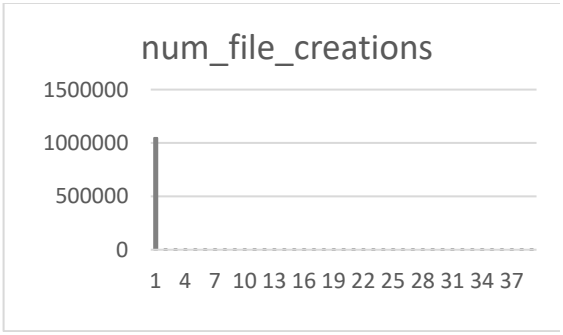


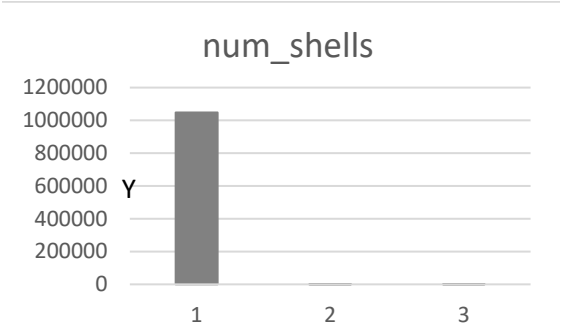
Figure 3. Analysis value of 10 Features of Dataset.


is_host_login
1200000
1000000
800000
600000
400000
200000
0
1
2


is_guest_login
1200000
1000000
800000
600000
400000
200000
0
1
2


Count
300000
250000
200000
150000
100000
50000
0
1
25
49
73
97
121
145
169
193
217
241
265
289
313


num_file_creations
1500000
1000000
500000
0
1 4 7 10 13 16 19 22 25 28 31 34 37


num_shells
1200000
1000000
800000
600000
400000
200000
0
Y
1
2
3

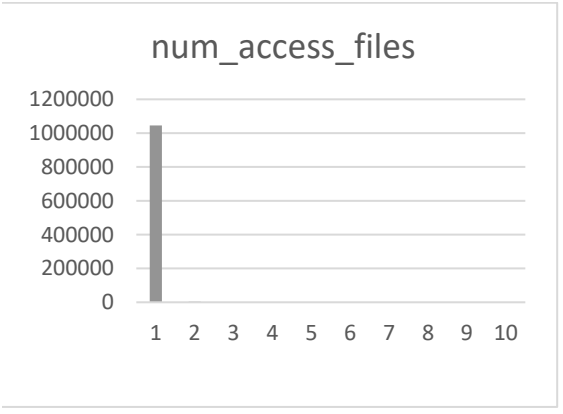
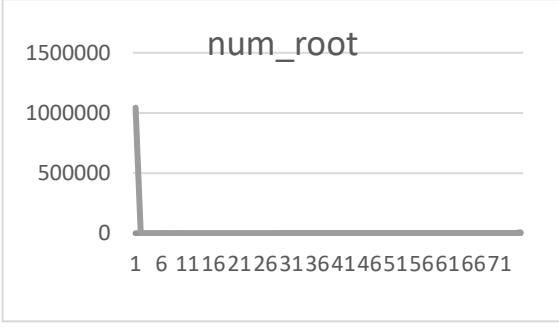

num_access_files
1200000
1000000
800000
600000
400000
200000
0
1 2 3 4 5 6 7 8 9 10

Table 1. Misuse Dataset 42 Features.

No. class	Feature name	Total number of attack
1	Root_ num	The root in the connection is done as a number of operations or the number of 'root' accesses
2	File_creations_num	The file creation operations were performed in the connection of the number
3	Shells_num	Shell prompts of the number
4	Files_access_num	Control the number of operations performed on accessing files.
5	Num_outbound_cmds	Commands that are sent out in an FTP session
6	Is_hot_login	If the login is on the 'hot' list, like root or admin, then it will be 1; otherwise, it will be 0
7	Is_guest_login	1 if the login is a “guest” login; 0 otherwise
8	Count	The number of connections completed in the last two seconds with the same host as the actual connection destination.
9	Srv_count	The number of connections made to the same service (port number) within the preceding 2 seconds as the current connection
10	Serror_rate	The proportion of connections among those aggregated in count that have triggered the flags s0, s1, s2, or s3 (23)

The analysis value of 10 Features of Dataset explain in figure 2



is also, where the suggested system is trained. The second half, which makes up 30% of the dataset, is utilized to test the suggested system. Then, assess the suggested system's accuracy in classifying the abuse feature algorithm, which is offered in multiple steps as follows: (Matrix Data Set as Input; Best Tree as Output)

1: in the first step the samples will be selected from the dataset.

2: After construction, a decision tree for all samples will provide a prediction result.

2.1 evaluate the previous entropy.

2.2: evaluate the information Expanded for the attribute.

2.3: evaluate the attribute value then partition the data set based on these values.

2.4: Repeat Steps 1- 3 per Tree using the relevant partition.

3: in the last step the voting for every expected result will be completed.

3.4.2 Testing Dataset

The KDD dataset contains approximately one hour of anonymized traffic traces from the misuse attack and Distributed Denial-of-Service (DDoS) attack. The server is able to connect to the internet and its computer capabilities by using all of the network bandwidth. This attack aims to prevent users from accessing the targeted server. The KDD dataset contains more than 4 million requests and connections. Table 1 shows that each connection had 10 fields of information. It is also the analysis value in Table 2. This dataset has 972,517 connections from misuse attacks, and the remaining belong to DDoS traffic and normal traffic.

highest value is a global "best" value and is referred to as g_{best} when a particle uses all the population as his topological neighbor. . The pseudo-code of the proposed PSO model algorithm is given in several steps, as follows:

Step1: find variance γ2 of the data by using standard deviation γ.

Step2: Determine the null and alternate hypothesis where N0: no difference in variances and N_a : difference in variances

Step3: Find F_{calc} using equation (1), $F_{calc} = \gamma_1^2 / \gamma_2^2$, where F_{calc} = Critical F-value, $\gamma_1^2 \& \gamma_2^2$ = difference of the two samples.

Step 4: Find the degrees of freedom of the two samples using equation (2), $d_f = n_s - 1$, where: d_f = Degrees of freedom of the sample, n_s = Sample size.

Step 5: Find F_{table} value using d1and d2, obtained in Step4 from the F-distribution table using learning rate a = 0.03, $d_1 = d_f$ of the bigger sample with numerator variance,$d_2 = d_f$ of the smaller sample with denominator variance.

Step 6: Interpret the results using F_{calc} and F_{table}. where Max referred to the best value the feature takes , Min referred to lowest value the feature takes ,range = $|Max - Min|$, D represents discrete values of feature ,and C represents continuous values of feature.

3.4 Classification Stage

The Proposed Misuse Attack Detection Based on Data mining in NFV

3.4.1 Random Force classification algorithm

The dataset was divided in the proposed model into two parts using the Random Force classifier. 70% of the dataset is in the first section, which

II. Data Cleaning

To complete the process, one must identify inaccurate data, eliminate unnecessary information from the data, and replace the incorrect data. The real data cleaning process is error elimination and data validation. Cross verification of data can help you find and correct errors. Validating the data is a way to resolve the problem.

III. Model Building

To describe the behavior of a variable, a statistical model or machine learning model is utilized. It is possible to have both supervised and unsupervised models. A classification or regression model can be used to obtain the results. The outcome can be seen by using a model. After that, the model must be evaluated.

IV. Present Result

By using charts, graphs, and tables, we may view large amounts of complex data. Graphs can help people digest information. The idea can be explained with a simple approach. The areas that need improvement can be identified by it. It effectively clarifies the component [24].

3.3 Feature selection Using PSO algorithm

The population of the particles in the first swarm is typically spread randomly over the search space. In every iteration, both "best" values, p_{best} and g_{best}, are used to update each particle. In the issue space, the coordinates of each particle are recorded that correspond to the best solution (fitness) it has found so far. The name p_{best} is given to this fitness value that is kept. The

3.1 Pre-processing Stage

Data preprocessing is an important stage that has been taken into consideration before the further processing stage. In this work, One Hot Encoding is used for dataset preprocessing. In order to convert categorical variables to numerical values in an interpretable format, the most well-known technique, one-hot encoding, has been, in addition to the job of quantifying categorical data in machine learning. Additionally, One Hot Encoding generates a vector whose length may equal the number of feature categories in the test dataset. Except for the ith component, which is given a value of 1, all components of this vector are assigned the value 0 if the data point falls into the ith category. Following categories logically in terms of numbers is made possible by this. However, it is quickly fixed by the statement that approximately one hot encoded has exactly one position in its array that is labeled as a (1).

3.2 Exploratory data analysis (EDA)

Techniques for data visualization and analysis are widely employed. Following is a discussion of these methods:

I. Data Exploration

That is the first step in analyzing the data. This is where we can learn more about the content and characteristics of the data set. The size of the data is exposed. Locating the missing value of the data is possible. We can see if there are any potential connections between the data. By using tabular data and comprehending its properties, data visualization is accomplished.

$$Accuracy = (\frac{TP + TN}{TP + TN + FP + FN}) \times 100 \quad \textbf{(4)}$$

TP and TN are synonymous with True Positive and True Negative, which suggests that the proportions of positive and negative conditions were properly identified. Also, FP means false positive, which refers to all the negative cases that were falsely classified as positive, and False Negative, or FN, is the term for all positive cases that were misclassified as negative [23].

3. The Proposed System

A machine-learning algorithm is using in the proposed system to achieve a higher accuracy rate for detecting misused attacks. The input into the system is a KDD data set, and the output is the attack rating. The proposed work consists of several main stages, namely loading the KDD dataset, preprocessing the encoded dataset using the One Hot Encoding technique, selecting the most important feature from the dataset using the PSO technique, and classifying the results using Random Forest and Niave Bayse. The diagram of the proposed model's general blocks is shown in Figure 2.

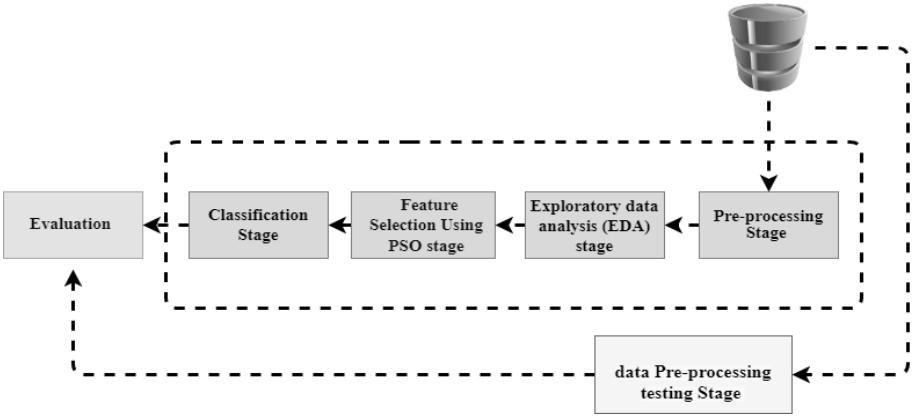


Figure 2. The Proposed Model's General Block Diagram.

making with KDD is to reinforce and enhance DSS through the models created for KDD, especially in situations where a significant amount of historical data is accessible for knowledge discovery [18].

2.4.2 Random Forest Classifier

All satellite scenes from each year were classified using the random forest method [19]. This method, which is a variation of the classification tree algorithm, is used in ensemble learning techniques. Based on tree-wise randomly selected samples and subsets of the training data, individual decision trees are generated automatically by the classification tree algorithm. Many classification trees are created for a random forest model, and the classification outcome is determined by a vote of each tree. A random sample of the training data set is used to create each classification tree individually using a unique learning method. The best split is carried out at each node of classification trees using random selections of the predictor variables. Every tree grows to its maximum potential under the control of the user-set node size. There are a number of significant adjustable factors that must be set before the random forest classification can be used. The main parameters are the depth of each tree in the forest, the number of classification trees to run in the forest (ntree), and the number of randomly chosen variables to utilize in each tree's construction (mtry) [20].

2.4.3 Evaluation Methods

A confusion matrix is a form of visualization method that is widely used to verify the accuracy of classifiers in classification [21]. Accuracy is a parameter that tests a classifier's capacity to render a correct diagnosis [22]. The accuracy of the equation is seen in equation 4.

rendered prior to the final classification. Although supervised classification is far more reliable than unsupervised classification, misclassifications can be identified because it is heavily reliant on teaching. The monitoring of classification involves careful attention to the creation of training details. Classification outcomes would be bad if training data were poor or not representative. As a consequence, controlled labeling usually takes more resources and time than unsupervised [16].

2.4.1 KDD Datasets

The capacity to select the best option from a range of choices to address a specific issue or accomplish a set of goals is known as decision-making. Strategic, tactical, and operational decisions may often be made in any organization or system. In any company, decision-making processes can be supported in a variety of ways. A well-known strategy employed in businesses is the decision support system (DSS). A decision support system (DSS) is an information system that supports the management, operating and planning levels of an enterprise, organization, or business and addresses a variety of issues to help decision-makers.

A well-designed DSS can assist decision-makers in gathering information from numerous sources and in making judgments by resolving issues [17]. Data mining is the main KDD sub process that uses various algorithms and techniques on databases to create predictive models and extract useful and instructive patterns. The maternal health and child immunization databases can be utilized to identify previously unidentified patterns and knowledge, which can then be used to create more effective and efficient decisions for the management of healthcare. The fundamental goal of integrating decision-

$$If\ (rand < sv_{pd}^{new}))\ then\ x_{pd}^{new} = 1; else\ x_{pd}^{new} = 0 \quad \textbf{(3)}$$

Where $v_{pd}^{new}\ and\ v_{pd}^{old}$ are the particle velocities, x_{pd}^{old} is the current particle position (solution), and x_{pd}^{new} is the updated particle position (solution). The values $pbest_{bd}\ and\ gbest_{d}$ defined as stated above. The two factors $rand_1\ and\ rand_2$ are random numbers between (0, 1).

2.4 Classification

The process of predicting the goal class of each data sample includes utilizing a supervised learning approach that determines the classes to be used beforehand. When utilizing classification algorithms, data attribute values must be utilized to determine class definitions [15]. The first stage in the classification phase is called preparation, which is necessary to construct a model for classifying new data. The model is used to forecast the class mark of the new data in the second step, which is referred to as classification. To describe a previously established set of data classes, a classifier is built during the training phase. The classification algorithm creates the classifier by acquiring knowledge from training data and related class labels. This phase can be regarded as a learning phase where the function y = f(x) can predict the associated class label for the given data, and the prediction's accuracy is the first thing to be calculated. If you use training data for computing accuracy, it will be optimistic (the classifier leads to overflow the data), so in this state, a test set is employed where the data is not dependent on the training set. If the accuracy of the classifier is acceptable when using the testing set, it can be used in the future for the data whose class label is unknown. In the controlled classification, the bulk of the initiative can be

Effective EDA in this area also heavily relies on the abilities and domain expertise of data scientists [15].

2.3 Feature Selection Using PSO

James Kennedy and Russell Eberhard created particle swarm optimization (PSO). It was created using a straightforward idea inspired by the movement of fish schools and bird flocks. It was created following several computer simulation-based interpretations. PSO uses several agents (particles) that come together to form swarms. In search of the best solution, this swarm is moving around in the search area. To match its own and other particles' flying experiences, it modifies each particle's "flying" in the search space.

PSO started with randomly created particles and their velocities, which represent the speed of the search. The particles are then assessed for fitness. Two major tests are conducted following this examination. The first test, known as personal best, compares a particle's experience with itself (pbest). The second test contrasts a particle's fitness with the overall swarm experience. The phrase "global best" (gbest) for the best particle is saved as a result of running these two tests. The termination criteria are then satisfied [16]. A binary string that corresponds to the feature selection case determines each particle's position. The formulas below determine the frequency of updating each particle (1, 2, and 3):

$$v_{pd}^{new} = w\,x\,v_{pd}^{old} + c_1\,x\,rand_1\,x\left(pbest_{bd} - x_{pd}^{oid}\right) + c_2\,x\,rand_2\,x\left(gbest_d - x_{pd}^{old}\right) \quad (1)$$

$$s\left(v_{pd}^{new}\right) = \frac{1}{1+e^{-v_{pd}^{new}}} \quad (2)$$

This is not the case in the case of NFV [12].The role of a Network virtualization is a cloud infrastructure networking implementation that virtualizes network services that can be chained together to construct a virtualized communication infrastructure. This allows a shared, flexible, and energy-efficient network ecosystem to be developed. TSPs are referred to as "NFV providers" in this situation. As shown in Figure 1, the European Telecommunications Standards Institute (ETSI) has established an NFV structure with three key components [12, 13].

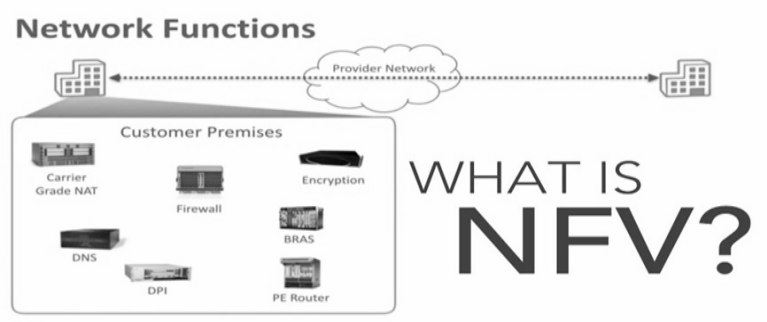


Figure 1. General architecture of NFV [13].

2.2 Exploratory data analysis (EDA)

The primary purpose of exploratory data analysis is to examine what the data can reveal outside of formal modelling or hypothesis-testing tasks. The analysis of data sets is assisted by EDA, which focuses on four essential elements to highlight their statistical characteristics: the distribution's shape, the presence of outliers, and the measures of central tendency (including the mean, mode, and median) are all taken into account, and the measures of spread (which include the standard deviation and variance). Data scientists evaluate the data and derive important insights using EDA techniques.

and virtualized networked environments. While coefficients and signal shifts were employed to complete period detection, SVM was used to identify cycle numbers.[10] Created an intrusion detection system by employing a recurrent neural network. The proposed mechanism has three stages: data collection, characteristic extraction, and the deep threat classifier. The results of the studies show that Naive Bayes, K-Nearest Neighbor, Random Forest and Decision Tree are ineffective compared to the best precision of 98.18% according to the 10-fold cross-validation analysis.

In [11], network intrusion detection (NID) systems have made extensive use of data-driven methodologies. The way the datasets are gathered, however, has led to several difficulties. In comparison to regular traffic, the majority of attack classes in network intrusion datasets are regarded as the minority, and many datasets are gathered using virtual machines or other simulated settings as opposed to real-world networks. By fitting models as random forests or support vector machines to non-representative "sandbox" datasets, these problems reduce the performance of intrusion detection machine learning models.

2. Related Theories

In the field of our research, we relied on several modern techniques and methods to build the proposed model which are listed as follows:

2.1 Network Function Virtualization (NFV)

Telecommunication service providers (TSPs) have historically been forced to implement physical patented equipment for each network purpose, resulting in high construction costs and constraints when scaling a network.

in the network environment by introducing frequent pattern mining in the detection process.

Niave Bayse, a machine learning technique, has been added to enhance the detection of misuse attacks. To test and evaluate the performance of the proposed model, the KDD dataset is utilized.

The remainder of this work divided into the following sections. Details on related theories are presented in Section 2. The suggested adaptive misuse attack detection model and associated techniques are described in Section 3. Experimental results are presented in Section 4. Section 5 ends our study by discussing the most important points of the system and comparing our proposed model with current methods

For the purpose of swift and verifiable identification of misconfigurations breaching security characteristics in NFV, [8] proposed a new methodology called MLFM which combines machine learning (ML) efficiency with rigorous formal methods (FM). The heart of our approach is an iterative teacher-learner interaction where the teacher (FM) can progressively (over several iterations) as training data, provide more representative verification findings, and the learner (ML) can use these results to gradually generate more accurate ML models.

In [9] covers the following goals: establishing and examining the two classifiers that are most frequently used in IDS implementation (KNN and Bayes), assessing the performance of each classifier separately, and modeling a hybrid classifier based on the advantages of the two classifiers. A quantitative methodology for data collection and analysis was used in this investigation. The NSL-KDD and the original KDD 1999 datasets were used in the investigation. Evaluated the developed techniques using traffic workloads

numerous studies have been established to tackle this type of attack, but no one has come up with the best solution.

The main point of this work is explained in the following steps:

1) Design and implementation of detection and diagnosis of misuse attacks in NFV networks with higher accuracy based on the Random Fours algorithm and Niave Bayse of machine learning.
2) The ideal categorization of normal and offensive behaviour in the intrusion detection model is achieved by combining NFV with a random forest and Niave Bayse classifier.
3) To improve the intrusion detection model, a frequent pattern mining-based adaptive online update strategy is added to the random forest and Niave Bayse classification and detection process is able to adapt to dynamic changes in the network environment and has a high detection rate for known and unidentified attacks.
4) The accuracy criteria are used to compare the performance of machine learning algorithms, knowing the best algorithm for diagnosing and distinguishing the misuse attack requires accuracy.

Our presentation presents an adaptive intrusion detection model that is based on NFV and the random forest algorithm to reduce the impact of dynamic changes in the network environment on the precision of the pattern for detecting abusive attacks, better extract effective data characteristics and enhance the detection performance of the abusive attack detection model. This work aims to propose a system capable of detecting early and accurate NFV misuse attacks using the most efficient machine learning technique the random forest .the intrusion detection model can adapt to dynamic changes

recognized as key technological paradigms in modern networking systems for efficiently handling emerging challenges in managing complex networks and there are unpredictable traffic patterns in various scenarios, including Internet of Things, cloud networking, and mobile data traffic in fifth generation (5G) and beyond networks [5]. Middle box functionality is progressively being virtualized and offered in software with the development of software-defined networking (SDN) and virtualisation of network functions (NFV), which can reduce energy consumption, resource use and service provider operating costs [6]. The typical traditional network is rigid and stiff. The Internet's rapid expansion has brought about challenges, including heterogeneity, scalability, and interoperability. A fundamental approach to virtualization in communication networks is Network Function Virtualization (NFV) [7]. Virtual networks that provide services through virtual parts of virtual machines are known as Network Function Virtualization (NFV). It is easy to implement and up-to-date. NFV's low costs are due to the sharing of resources. Attacks are not uncommon for NFV, just like other networks. Misuse attacks are the most common attack in NFVs because all parts share the same resources, the attack's use of one or more resources that affect all parts of the NFV is particularly noteworthy.

A neural network-based detection system was proposed by them for botnets in NFV. The system was formed using data sets that were collected through traditional networks. Based on the study, their proposed detection system could detect up to 99% of botnet attacks. Various types of attacks can target Network Function Virtualization (NFV) in different forms and destinations. Network Function Virtualization (NFV) could be targeted by various types of attacks in different forms and destinations. Moreover,

1. Introduction

The most critical part that must be considered when developing systems, especially those that must connect to the internet, is information security. The identity risk management sector is where it belongs [1]. The usual approach is to prevent or reduce the risk of improper data usage, leakage, destruction, deletion, corruption, alteration, review, tracking, or devaluation. The main goal of Information management is to maintain data's as the confidentiality, integrity and availability, and or (CIA triad) supporting the efficiency of business [2]. Over the past decade, massive cyber-attacks against major organizations have multiplied. Among the most devastating are ransomware, which aims at encrypting the data stored on the system until the target organization pays a ransom.

At times, security threats are on the rise. A computer system that has a flaw or weakness, security tactics, internal controls, and planning, A framework's security policy can be compromised by a web vulnerability, which could lead to data theft or modification. SQL injection and cross-site scripting (XSS) attacks, as well as broken authentication and session management, are the most common types of attacks. Weak or broken authentication mechanisms may allow attackers to circumvent authentication and gain unauthorised access to the application [3]. Resources on the NFV network are vulnerable to abuse, particularly those that only a small number of users or cannot share. In these attacks, the attacker takes resources away from other users and uses them for their purposes without sharing them with other users. These attacks typically cause bottleneck issues, which cause service delays or even the interruption of NFV network services [4]. Network Function Virtualization (NFV) and Software Defined Networking (SDN) are

المستخلص

هجوم سوء الاستخدام هو النوع الأكثر شيوعًا وخطورة من الهجمات التي يمكن أن تستهدف NFV المحاكاة الافتراضية لوظائف الشبكة . في هجوم سوء الاستخدام، يحاول المهاجم استهلاك موارد بيئة NVF عن طريق إرسال كمية كبيرة من حركة المرور. لحماية بيئة NFV، تم اقتراح نظام مبكر ودقيق للكشف عن هجمات سوء الاستخدام استنادًا إلى NFV و Random Forest Classifier. يبدأ النموذج المقترح باستيراد مجموعة البيانات وتحليلها ثم المعالجة المسبقة واختيار الميزات باستخدام خوارزمية اختبار PSO (تحسين سرب الجسيمي)، ويعتمد التصنيف على تقنية التعلم الآلي الأكثر شيوعًا وكفاءة، وهي Random Forest وNiave Bayse المستخدمة. وأخيرا، من أجل اختبار وتقييم أداء النموذج المقترح، تم استخدام مجموعة بيانات KDD (اكتشاف المعرفة في قواعد البيانات). تفوق النظام المقترح على الأبحاث السابقة الأكثر مقارنة من حيث الأداء، كما أشارت النتائج، حيث حقق معدل دقة 99.98% لـ Random Forest و 99.5% لـ Niave Bayse .

الكلمات المفتاحية : شبكة NFV, التعلم الالي، تصنيف الغابة العشوائي

Abstract

A misuse attack is the most common and dangerous type of attack that could target NFV (Network Function Virtualization). In a misuse attack, the attacker attempts to consume the NVF environment's resources by sending a large amount of traffic. To protect the NFV environment, an early and accurate misuse attack detection system has been proposed based on NFV and Random Forest Classifier. The proposed model starts with importing the dataset and analyzing it then pre-processing and feature selection using a PSO (particle swarm optimization) Test algorithm, classification is based on the most common and efficient machine learning technique, which is Random Forest and Niave Bayse used. Finally, in order to test and evaluate the performance of the proposed model, the KDD (knowledge discovery in databases) dataset is utilized. The proposed system outperformed the most comparable previous research in terms of performance, as indicated by the results, achieving an accuracy rate of 99.98% for Random Forest and 99.5% for Niave Bayse.

Keywords: Network Function Virtualization (NFV), Machine Learning ML, Random Forest (RF), Niave Bayse, PSO Test algorithm, KDD.

Protection the NFV Net-work Using the Random Forest Classification/ Original article

Mustafa H. Taha*[1],

Ibtesam Jomaa Ibrahem[2],

Wafaa Waheeb Abdullah Ali[3]

1 Department of Human Resources, General Directorate of Diyala Education, Diyala, Iraq,

2 Department of Computer Science, Diyala University President, Diyala, Iraq,

3 Department of Planning, General Directorate of Diyala Education, Diyala, Iraq,

حماية شبكة (NFV) باستخدام تصنيف الغابة العشوائية

مصطفى حسين طه[1]*،

ابتسام جمعة ابراهيم[2]،

وفاء وهيب عبدالله[3]

1 قسم الموارد البشرية, المديرية العامة لتربية ديالى, ديالى, العراق

2 رئاسة جامعة ديالى, جامعة ديالى, ديالى, العراق

3 قسم التخطيط, المديرية العامة لتربية ديالى, ديالى, العراق

* mustafahusseintaha@gmail.com

[4] Ohood Fadil Alwan, (2023), "Using Artificial Intelligence to Diagnose Demented in the Elderly, Al-Esraa University College Journal for Engineering," V (5)-No. (8), pp 107 – 140.

[5] S. Erol, A. Jäger, P. Hold, K. Ott, W. Sihn, (2016) Tangible Industry 4.0: a scenario-based approach to learning for the future of production, 6th CLF - 6th CIRP Conference on Learning Factories, Procedia CIRP 54 pp13 – 18.

[6] Callaway, E. H. (2004), "Wireless sensor networks: Architectures and protocols". CRC Press.

[7] ZigBee Alliance. (2007), "ZigBee Specification" (Document 053474r17). Retrieved from https://zigbeealliance.org/

[8] Saurabh Vaidya, Prashant Ambad, Santosh Bhosle, (201 8), "Industry 4.0 – A Glimpse ", Procedia Manufacturing issue-20, pp 233–238

[9] Nazmus S. Nafi, Khandakar Ahmed, Manoj Datta, Mark A. Gregory (2014), "A novel Zigbee based pilot protection scheme for smart distribution grid", Conference: Telecommunication Networks and Applications Conference (ATNAC), Southbank, Australia DOI:10.1109/ATNAC.2014.7020889

[10] F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, (2002), "A survey on sensor networks," IEEE Communications Magazine, vol. 40, no. 8, pp. 102-114,.

[11] G. Barrenetxea, F. Ingelrest, G. Schaefer, M. Vetterli, O. Couach, and M.Parlange, (2008), "SensorScope: out-of-the-box environmental monitoring" in Proc. 7th International Conference in Information Processing in Sensor Networks, pp. 332-343,.

[12] R. A. Saeed, S. Khatun, B. M. Ali, and M. K. Abdullah, (2009), "Performance enhancement of UWB power control using ranging and narrowband interference mitigation," in *Proc. Broadcom Corp. Conf. on UWB*,.

[13] Ankur Tumar, (2011), "Introduction to Zigbee Technology", Global Technology Center, Volume1,.

[14] The Verge. (2025), "The Tech That Could Turn Millions of Zigbee Light Bulbs into Motion Sensors with a Single Update, https://www.theverge.com/2025/1/22/24348688/zigbee-ambient-sensing-philips-hue-ivani-sensify

of data is determined by a number of factors, including collisions, obstacles between nodes, and network topology.

- the productivity in the first scenario is greater than in the second and third scenarios, and the productivity in the fourth scenario is very low because there are not enough coordinators to receive all the packets, meaning that the productivity in the fifth scenario will be very high because there is a large number of coordinators, as shown in the figures.
- Traffic sent parameter is a large in the first scenario due to the short time in generating packets and is larger in the fourth scenario due to the large number of nodes and routers.
- traffic received number is low in the fourth scenario, which contains a large number of nodes and routers, and this number is almost the same for the second, third and fifth scenarios, and has the highest value in the first scenario, meaning that almost all the packets sent are received.

References

[1] Scarlet A (2001), “Integrated control of agricultural tractors and implements: a review of potential opportunities relating to cultivation and crop establishment machinery”, in Journal of Computers and Electronics in Agriculture, Val. 30, PP.167□191.

[2] N.Wang, N.Q.Zhang,M.H..Wang, (2006), “Wireless sensors in agriculture and food industry recently development and future perspective”, Journal of Computers and Electronics in Agriculture, Val. 50 (1):1-14.

[3] R. A. Saeed, H. Mohamad, and B. M. Ali (2009), "Dynamic hybrid automatic repeat request (DHARQ) for WiMAX - mobile multihop relay using adaptive power control," *Comput. Commun.*, vol. 32, no. 5, pp. 806-813.

Conclusion

By performing the simulation and displaying the results for a several network (OPNET, and to reach the best possible design of network According to the necessary network parameters. Time delay, loss of packets, throughput, traffic sent, traffic received and productivity, we clarify:

- The results obtained in this paper were used in practical applications in the field of monitoring and Controlling greenhouses.
- That in the first, second and third scenarios, the time delay between the final nodes is not affected by the number of nodes when it does not exceed 20 nodes the time delay is the same. That is affected by the density of data packet flow across the network, but it increases in both the fourth and fifth scenarios, that is, when the number of nodes in the network reaches 240 nodes, due to the large number of nodes that make up the network.
- We note in particular The time delay is greater in the fourth scenario due to the presence of one coordinator device, while in the fifth scenario there are (13) coordinators.
- The number of discarded packets decreased in the second and third scenarios compared to the first scenario when the time between packets generation changed, which led to a reduction in the density of packets in the network. This loss in the fourth and fifth scenarios was similar, meaning that the presence of routers did not affect the operation of the network. The following figure shows this:
- The increase in the number of users of the wireless medium is the reason behind the increase in interference, and thus the amount

of the decrease in data packets generated primarily within the network and thus the packets sent.

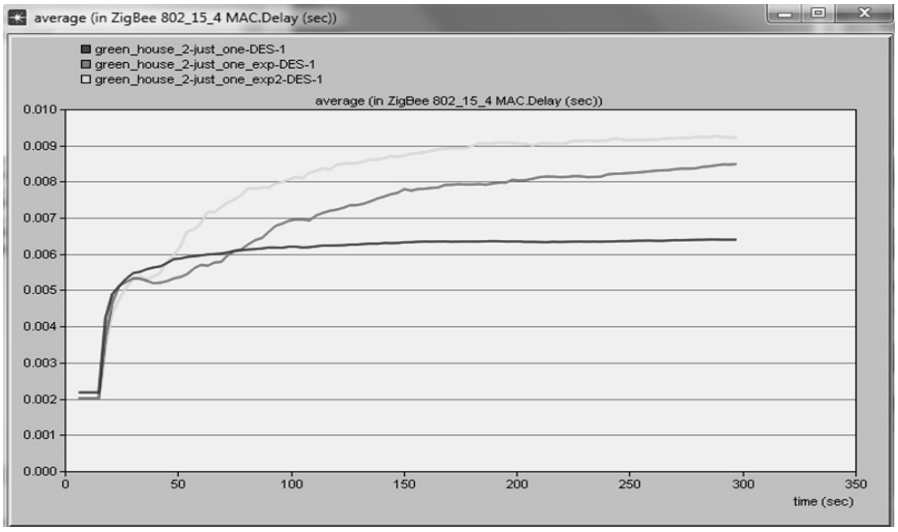


Figure (13) MAC DELAY

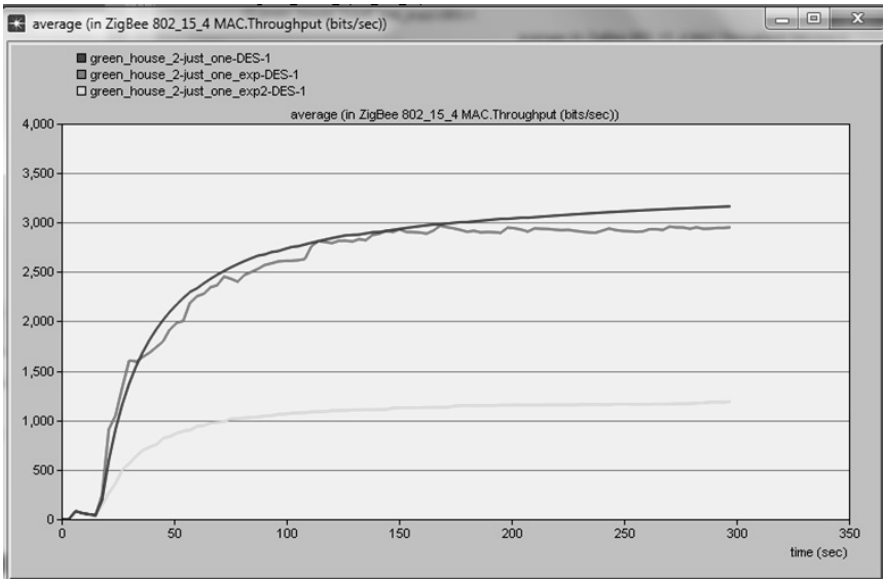


Figure (14) MAC-throughput

We notice that this number is low in the fourth scenario, which contains a large number of nodes and routers, and this number is almost the same for the second, third and fifth scenarios, and has the highest value in the first scenario, meaning that almost all the packets sent are received. As it illustrated in figure (12)

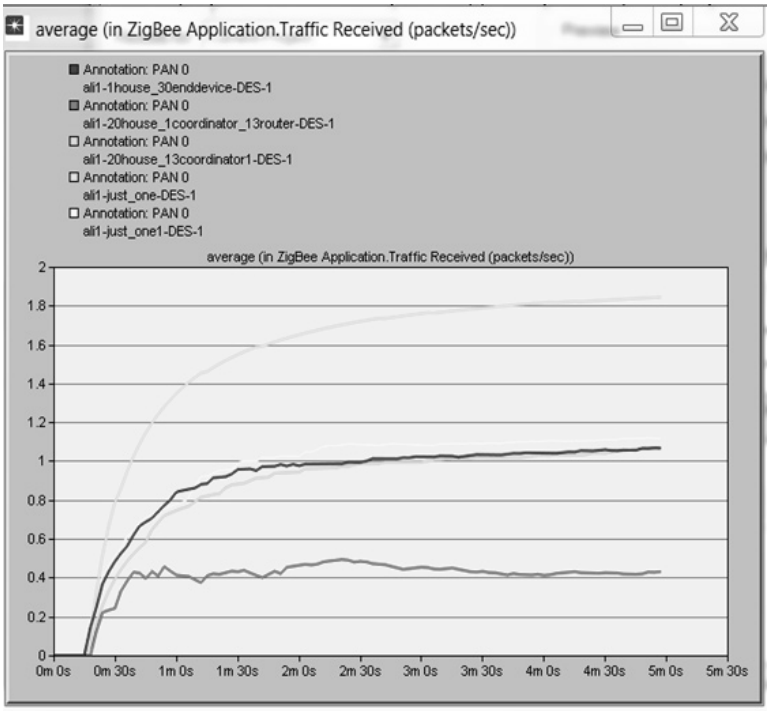


Figure (12) shows the graph of the transmitted and received data.

- **Delay:**

It is the time between waiting for the packet to be transmitted at the physical layer and receiving the last bit at the receiving node.

After conducting the simulation and displaying the graphs of productivity and delay, we notice an increase in delay and a decrease in productivity in the third scenario compared to the second scenario as a result

- **Data sent and received traffic sent & traffic received**):

- *traffic sent*: It represents the total number of data sent from the source to the destination in one unit of time regardless of whether it arrives or not.

We notice from the graph of this parameter after running the simulation as shown in figure (11), that it is large in the first scenario due to the short time in generating packets and is larger in the fourth scenario due to the large number of nodes and routers.

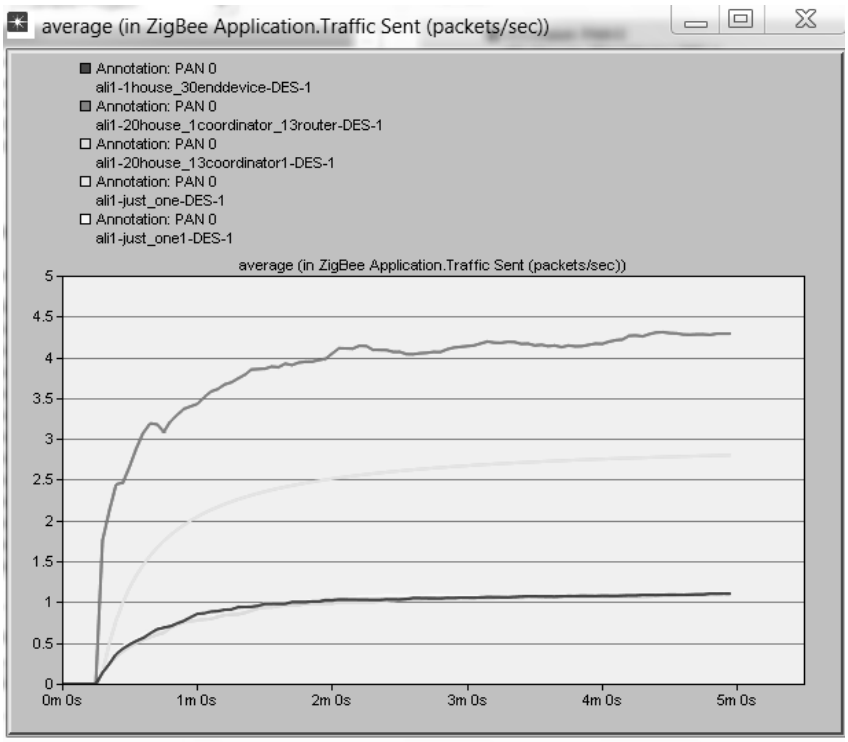


Figure (11) Chart of sent packets

- *traffic received*: Represents the number of packets received in one unit of time.

- **Throughput**:

The actual amount of data sent from the source to the target during a specified time (Second) The increase in the number of users of the wireless medium is the reason behind the increase in interference, and thus the amount of data is determined by a number of factors, including collisions, obstacles between nodes, and network topology.

After conducting the simulation and displaying the productivity charts for all scenarios, it became clear that the productivity in the first scenario is greater than in the second and third scenarios, and the productivity in the fourth scenario is very low because there are not enough coordinators to receive all the packets, meaning that the productivity in the fifth scenario will be very high because there is a large number of coordinators, as shown in the figure (10):

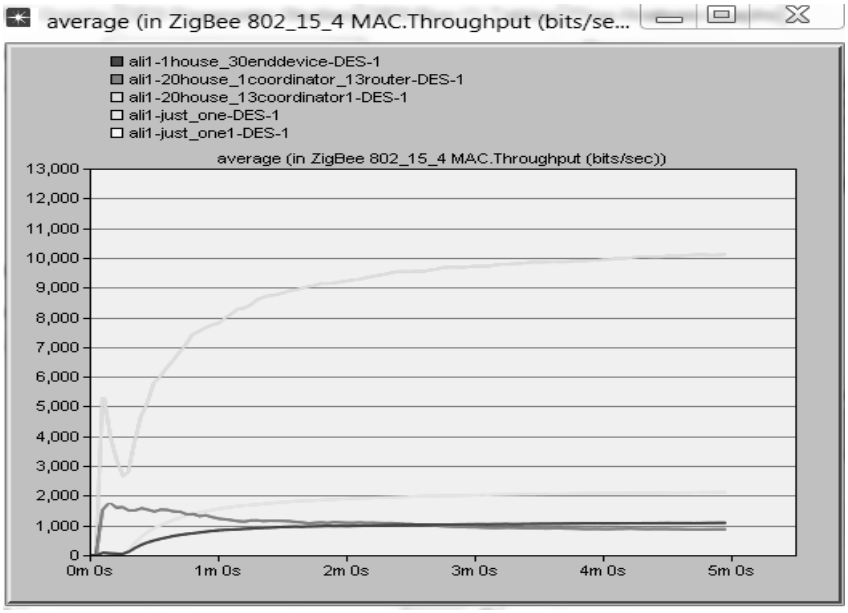


Figure (10) Productivity chart (throughput)

in the flow of data packets in it, so that the intermediate nodes' stores become full and unable to receive any new packets.

After running the simulation and displaying the graph of the discarded packets, it became clear that the number of discarded packets decreased in the second and third scenarios compared to the first scenario when the time between packet generation changed, which led to a reduction in the density of packets in the network. This loss in the fourth and fifth scenarios was similar, meaning that the presence of routers did not affect the operation of the network. The following figure shows this:

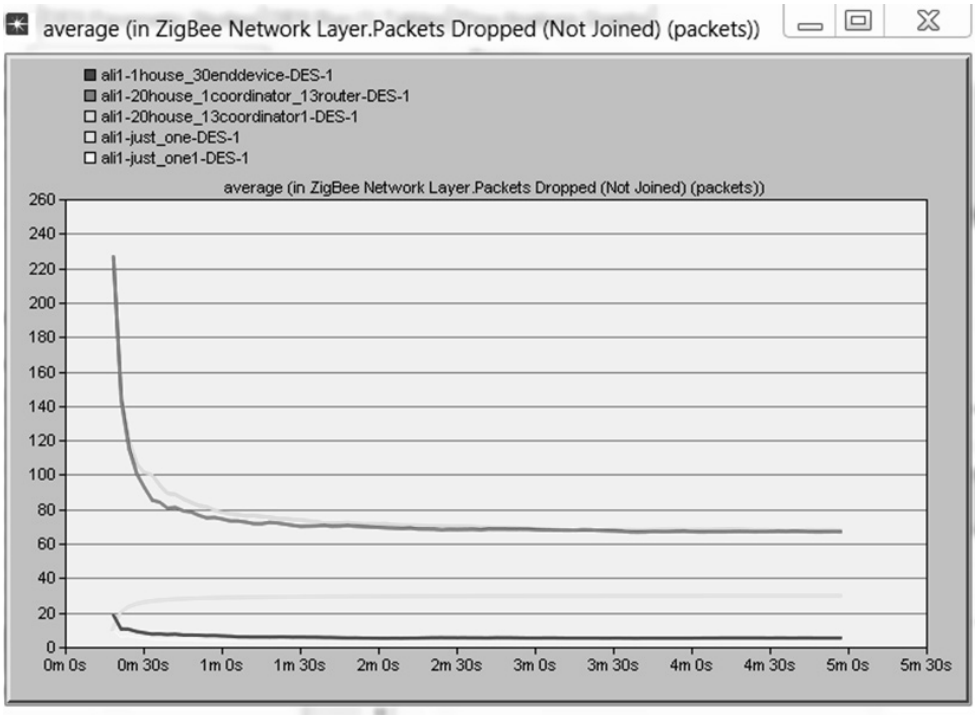


Figure (9) The graph of the neglected packets (loss of packets

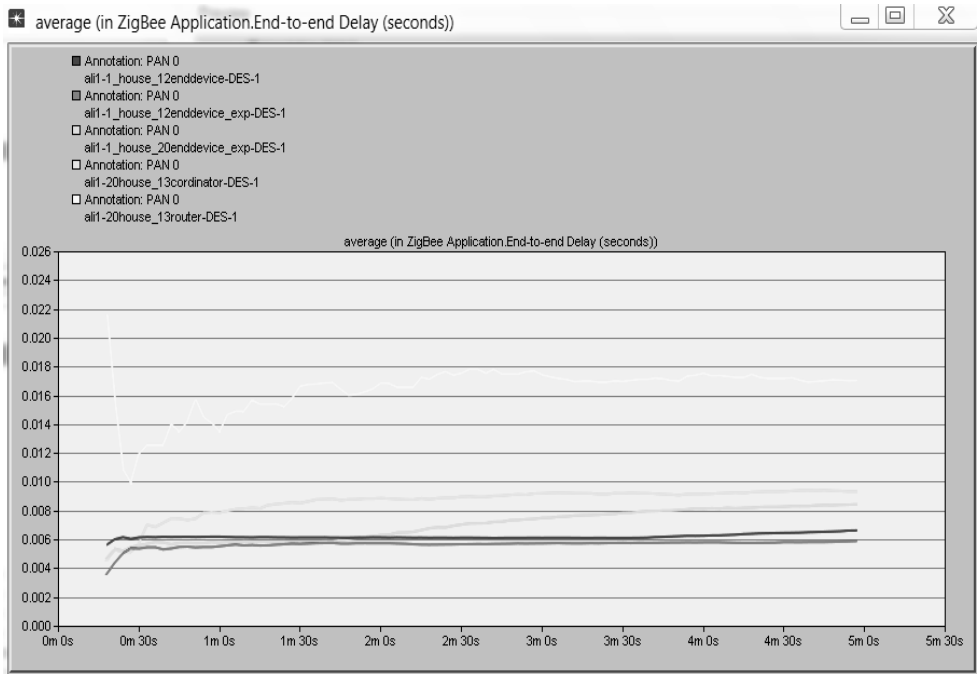


Figure (8) Time delay (end-to-end delay

- **Neglected packets (loss of packets**):

This parameter is defined as the parameter that determines the data packets sent from the sending nodes that do not reach the receiving nodes as a result of their loss in the network or the packets that do not reach from the sender to the receiver as shown in Fig-(8). There are several factors that cause packets to be lost while they are traveling within the network, such as: the absence of any receiving node within the range through which the sending node broadcasts its signals, which may decrease in value.RSS is a result of various losses or the end of the battery life of neighboring nodes located within its radio range, and one of the other most important reasons is the occurrence of a congestion condition in the network due to the increase

- **Time delay (end-to-end delay**):

It is defined as the total time delay between the sending node and the receiving node. By performing the simulation and displaying the results as shown in Fig-(8), it was observed that in the first, second and third scenarios, the time delay is the same.That is, the time delay between the final nodes is not affected by the number of nodes when it does not exceed 20 nodes, nor is it affected by the density of data packet flow across the network, but it increases in both the fourth and fifth scenarios, that is, when the number of nodes in the network reaches 240 nodes, due to the large number of nodes that make up the network. We note in particular: The time delay is greater in the fourth scenario due to the presence of one coordinator device, while in the fifth scenario there are (13) coordinators. The coordinator device represents the final node where all communications must end, and therefore the nodes located far from it will need a longer path and a greater number of hops, and thus a greater time delay to reach it, in addition to congestion and competition over the shared medium. As for the presence of more than one coordinator to enter the network, the nodes send data to the closest coordinator to the medium within the network, noting that the routing protocols are responsible for this task to form the parent-child tree or build the approved routing tables and a cage for the protocol and algorithm used, which is the same in all scenarios.

The following figure illustrates the characteristics of the proposed scenarios:

Inter arrival time	Size of data packets generated at the application layer	node spacing	Number of General Coordinator Contracts	Number of final nodes	Number and size of sites served	Scenario number
1second	1024	40	1	12	1 (200*160)	1
Exponentially increasing and step (10)	Exponentially increasing and step (100)	40	1	12	1 (200*160)	2
Exponentially increasing and step (10)	Exponentially increasing and step (100)	35	1	20	1 (200*160)	3
Exponentially increasing and step (10)	Exponentially increasing and step (100)	35	1	240	20 (200*160)	4
Exponentially increasing and step (10)	Exponentially increasing and step (100)	35	13	240	20 (200*160)	5
Exponentially increasing and step (10)	Exponentially increasing and step (100)	35	13	220	20 (200*160)	6

9. Simulation and results

After we have presented the design of this project with several network configurations and different parameters for the devices used, we will simulate these configurations using the chosen program (OPNET) and display the results in the form of graphs and analyze and discuss these results to reach the best possible design. As we mentioned in the introduction to this chapter, the comparison will be made on several parameters.

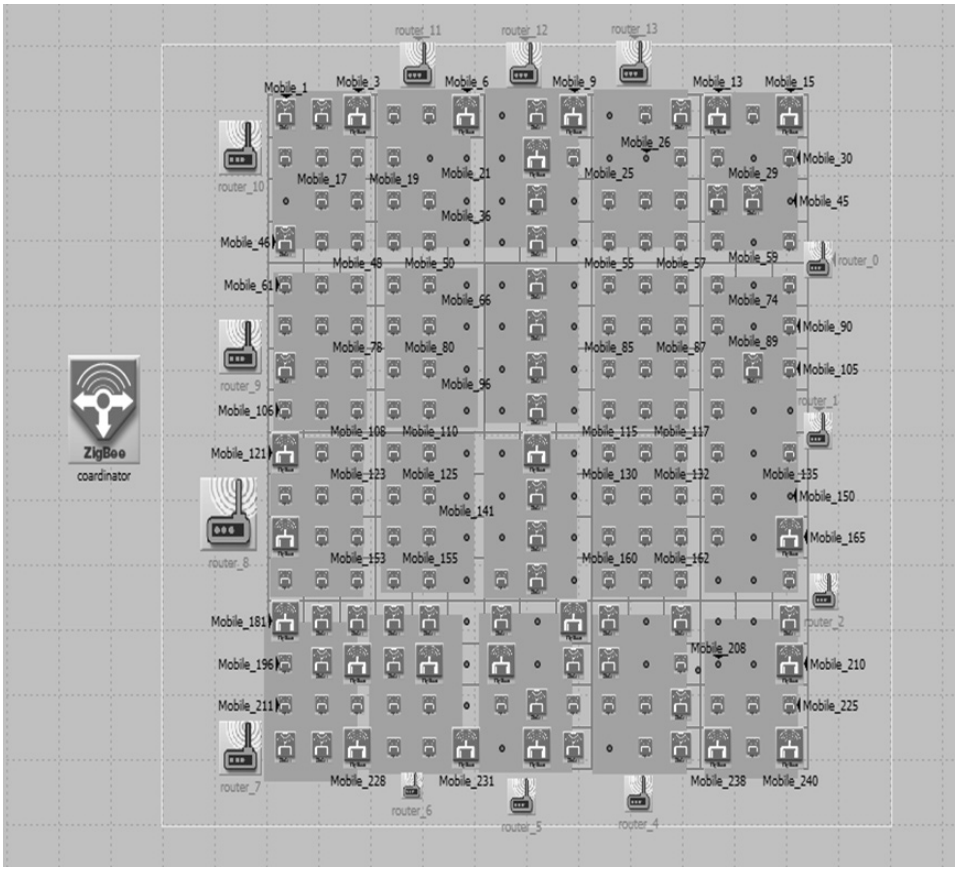


Figure (7) Network design illustrated for the fourth scenario

number of devices composing the network, as we will increase the number to 20 devices instead of 12 devices and we will make the distance between each two consecutive devices (30) meters..

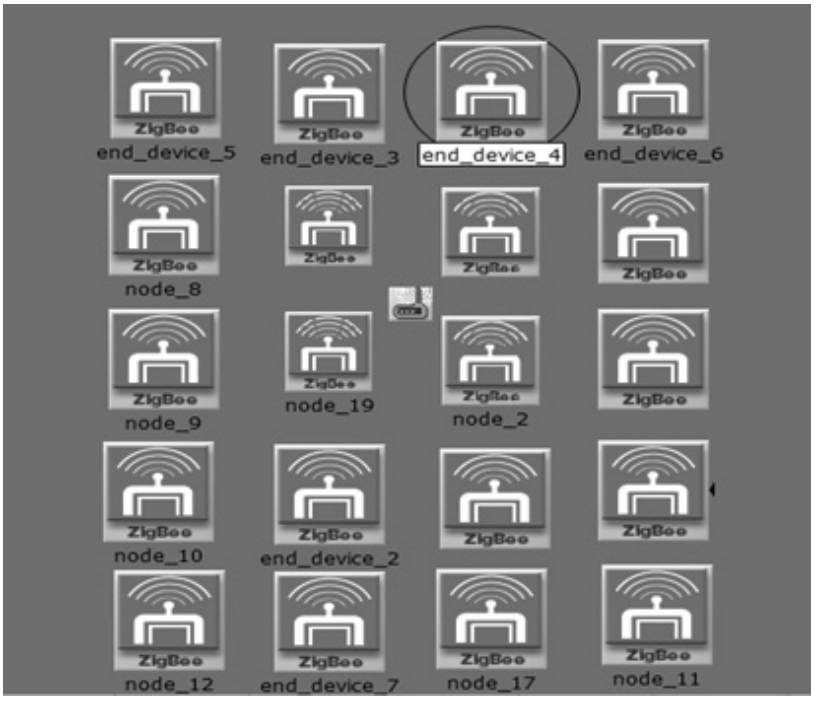


Figure (6) Network design in the third scenario

Scenario 4:

In this scenario, we will expand the network and increase the number of protected areas to become (20), each of which is identical to the design studied in the second scenario in terms of dimensions, area, number of nodes and their parameters. That is, this design will contain (240) end devices. However, we will use (13) routers here, and we will also use only one coordinating device for the design as a whole, as shown in the following figure (7):

Figure (5) first scenario, 12 Zigbee Nodes with one Coordinator

Scenario 2:

In this scenario, we will adopt the same design as in the first scenario in terms of dimensions, number of devices, and their distribution, but we will change the settings of the deployed devices. We will make the size of the transmitted data variable exponentially and with a step of (100). We will also change the time between generating two consecutive packets and make it variable exponentially as well and with a step of (10), as shown in figure-(6).

Scenario 3:

In this scenario, we will adopt the same previous design in terms of dimensions, the presence of one coordinating device and the same previous adjustment for device settings (the same size of the transmitted data and the same time between generating packets), but the change will be in the

- Productivity (throughput).
- Data sent and received (traffic sent & traffic received).

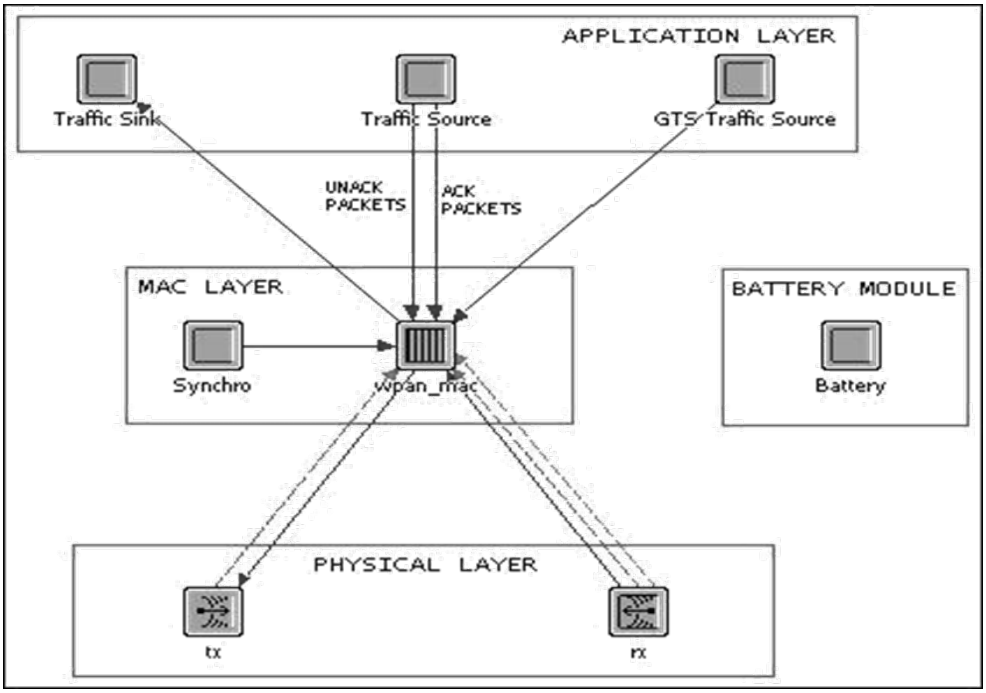


Figure (4) layers structure of Zigbee Protocol in OPENT

Scenario 1:

In this scenario as shown in Fig-(5) we have set up a protected area of dimensions (200*160) meters containing one coordinating device located in the center. The area has (12) final nodes representing the sensors, with a distance of (40) meters between each two nodes.

The default parameters of the devices used in the network were used in this scenario because they fit this design. The access method between nodes was set to random, meaning that each node sends to the closest node until the collected data reaches the primary node (coordinator).

The third model implements the media access protocol at the data layer.MAC.

The fourth model is responsible for routing data traffic, joining subnets to the overall network, and generating beacon requests.

7. Simulation using the program OPNET for IEEE802.15.4

The programOPNET contains version (1.0) of the accurate simulation model and this version is specific to and programmed for the IEEE802.15.4 standard. The PHY layer contains a transmitter and receiver operating at 2.4 GHz frequency, we use QPSK modulation. [12].

Contains a layer the MAC on Slotted CSMA-CA generates beacon frames and ensures synchronization of nodes with the network coordinator. We have the battery module which determines the power levels used. We also have the application layer which contains the data generator in the form of frames. The command frame generator in the MAC generates the ACK frames and the assembly module with statistics for the incoming frames. The radio model contains standard radio modules to simulate the radio channel along with other elements such as interference, noise, propagation delay.

8. Design and Simulation

In this research, we will study several scenarios with several differences, simulate them, compare the results, and analyze them to reach the best representation. There are many parameters that can be compared between different designs and in this research, we will rely on the study of

- Time delay (end-to-end delay).
- Neglected packets (loss of packets).

3. Transfer data between devices:

In style (beacon) cannot send between them.

In style ((non-beacon) Data is transmitted directly using un slotted CSMA/CA technology to access the channel.

The model is ZigBee in OPNET uses four operational models:

1. Model ZigBee MAC.
2. Model ZigBee Application.
3. Multi access carrier model ZigBee CSMA-CA.
4. Model ZigBee Network.

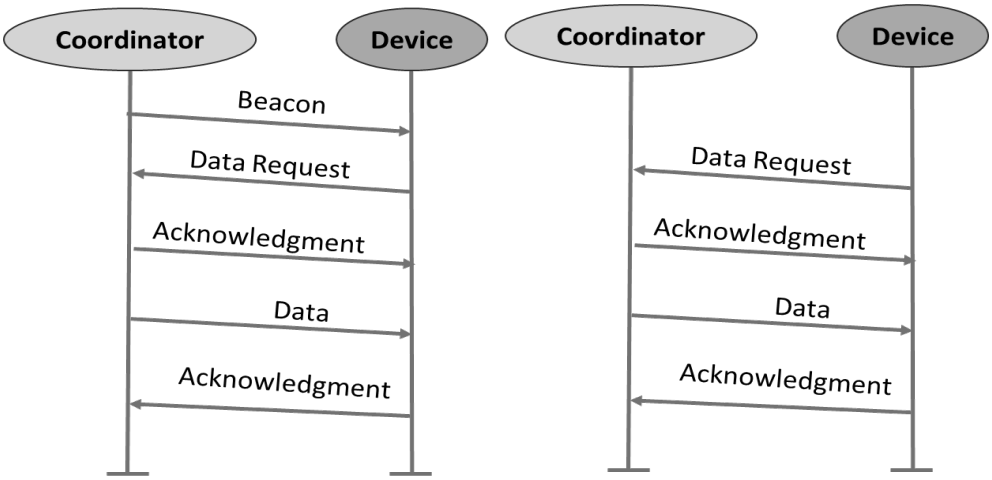


Figure (3) data transfer from the coordinator in Zigbee Protocol

The first model contains a model of (IEEE802.15.4 MAC Protocol) which is used for channel scanning, joining, failure and repair protocol in CSMA-CA unslotted system.

The second model represents a low-resolution version of the application layer in ZigBee contains network configurations, generates and receives traffic, and provides simulation reports.

The following figure-(2) shows the transfer of data to the coordinator:

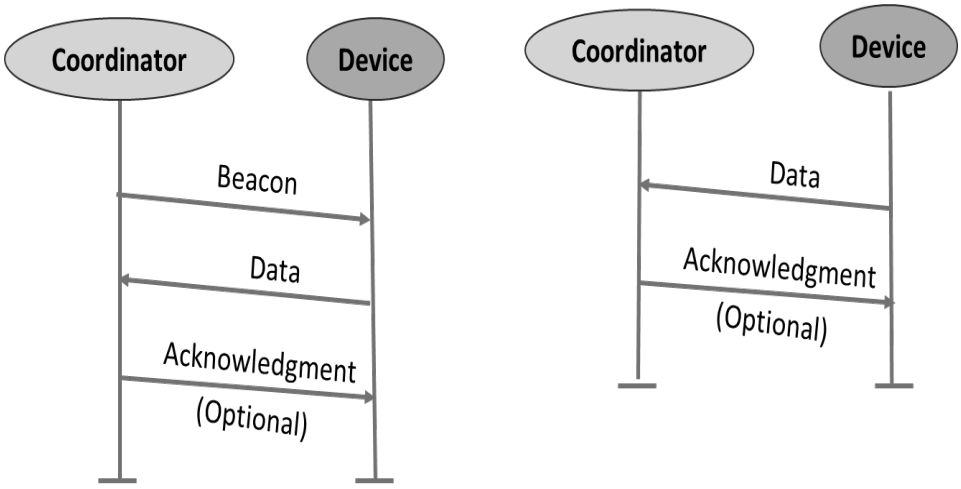


Figure (2) Data Transfer to the Coordinator in Zigbee Protocol Adapted [6]

2. Transfer data from the coordinator:

This is based on the coordinator's request, as it is in the indicative pattern (beacon) The coordinator tells devices that it has packets stored instead of directly sending frames to devices. The device that receives the beacon message first checks whether its identifier appears in the fields of the message. If it does, the device sends a request to send data to the coordinator. The coordinator responds with an ACK and passes the data to it.

As for the non-indicative pattern ((non beacon) The device must periodically send data request frames to inquire if there are packets stored for it, and if there are, the coordinator responds with an ACK and sends the frame to begin the data transfer procedures. The following figure-(3) shows the data transfer from the coordinator:

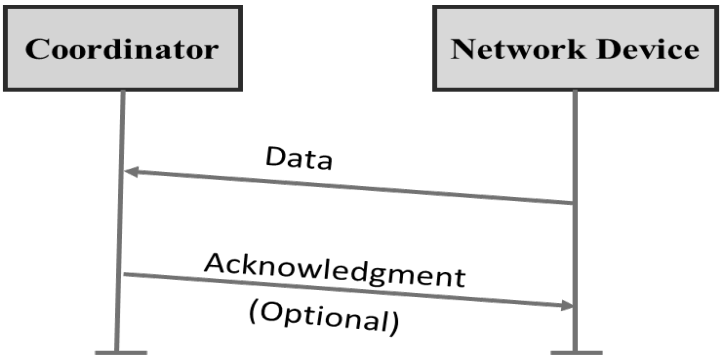


Figure (1) non-indicative pattern of Zigbee Protocol Adapted [6]

In general, the protocol ZIGBEE reduces the time that the sensor node is in the ON state and thus reduces the amount of energy consumed. Whereas in BEACON mode the sensor is only active during the data transmission period, while in NON-BEACON mode the energy consumption is absolutely not synchronous with the transmission since some sensor nodes are always in the ON state without working, so we notice that BEACON mode is the most important mode in WSN networks. So, we have three models for data transfer:

1. Transfer data to the coordinator:

Where in pilot mode when a particular device wants to transmit it uses time slots to compete for the channel using the technology of slotted CSMA/CA after receiving the beacon message.

In the non-guide mode, devices will compete for the channel using theunslotted CSMA/CA The data is sent directly to the coordinator and when the coordinator receives this data, it sends a report of it with an (ACK) message.

- **The guide pattern (beacon mode**):

In which devices are used the super frame that starts with the beacon helps in synchronization between the devices participating in the network. Each super frame is divided into an active period and a variable inactive period. Communication between the devices takes place during the active period and stops during the inactive period, and the devices enter a sleep state and low power mode until the beginning of the new frame.

Here the devices are used. Slotted CSMA-CA. This mode is used when the CORDINATOR is powered by battery. In this mode, the END DEVICES wait for a BEACON message from the CORDINATOR, which is broadcasted periodically, to send the data. After completion, the CORDINATOT and END DEVICES enter the SLEEP mode.

- **Non-indicative pattern (non-beacon mode**):

This pattern as shown in Fig (1) is employed when the feed is CORDINATOR via a power supply and in this mode the END is always in SLEEP mode and is activated when the required event occurs. For example, in wireless sensor networks inside the GREENHOUSE, the sensor nodes will wake up and become in an ACTIVE state when the temperature reaches the required value, then they send information to the CORDINATOR which remains in a continuous waiting state.

The disadvantage of this pattern is that the sensor node may find the channel busy and information will be lost.

This layer provides three important services:

Receiver Energy Detection (ED): It estimates the strength of the received signal within the channel bandwidth in order to determine the reception of this signal as a packet and is also used in the network layer to create a path.

The Link Quality Indication (LQI): is a process of characterizing the quality of the received packet and this characterization is done using an (ED) device in order to calculate the signal-to-noise ratio.

Channel Assessment (CCA): It is a logical function that exists to determine the current state of the wireless medium in order to be able to use it. In this standard, three types of CCA are used:

- **Energy over threshold**: Where it gives the CCA indicates that the channel is busy if the ED value is above the threshold.
- **Carrier Sense Only**: Gives the CCA states that the medium is busy only when a modulated signal is detected and has the propagation characteristics of the IEEE802.15.4 standard.
- **Carrier sensing with energy above threshold**: Gives the CCA indicates that the medium is occupied when a signal is detected with the IEEE802.15.4 standard modulation and propagation characteristics with power above the ED threshold.

6. The standard IEEE802.15.4 MAC ZigBee

It is located above the physical layer and works to secure access to the wireless channel by avoiding potential collisions in the channel. It provides services to the network layer and contains address data. frame to know the source and destination of this data, this standard allows two modes of operation in the network [9-16]:

- Equipment layer ZigBee is responsible for establishing local links between the source node and the target node, routing messages between devices, and defining device rules within the network such as: ZigBee Coordinator, Router or End Device.[14]

5-2 Network layer:

This layer is responsible for: - Initiating and defining the network topology (joining or leaving the existing network).

- Directing the Frame to target.
- Path discovery between peripheral devices (Sensors).

5-3 The standard IEEE 802.15.4 PHY Zigbee (Physical Layer):

It is the lowest layer and performs the process of modifying the outgoing signals and extracting the incoming signals. It also allows for channel selection to avoid radio interference, in addition to exchanging data with the higher layer (MAC)[7-13]. The IEEE 802.15.4 PHY standard at the physical layer supports two types of services:

- **Data Services**: Allows the transmission and reception of physical layer protocol data units. (PPDU: Physical Protocol Data Unit) over dedicated radio channels.
- **Management services**: These are related to sending pilot frames by nodes programmed as multi-access coordinators, and they have priority to start sending these frames using a timer that ensures the accuracy of the transmission timing.

formations (star, tree, mesh, etc.). It is also considered a low-cost technology compared to other technologies, in addition to its high, accurate and flexible performance.

Wireless sensor networks designed using sensors such as ZigBee provides high flexibility and diverse capabilities in all scientific and engineering fields. Using these networks, information can be transferred from one area to another at a low cost and with low energy consumption. The advantages of choosing ZigBee include:

- It expands the industry and processing systems and enables constant supervision and control that increases management and thus labor productivity.
- Supervising the work and thus saving personal effort and public safety.
- Get accurate readings, do precise work, and detect problems before they happen.
- Eliminates the need for manual monitoring and thus reduces the occurrence of unwanted errors.

5. Protocol structure of ZigBee

The technology ofZigBee has its own four-layer model, each with its own function, as in the OSI model. It includes:

5-1 Application layer:

Application layer ZigBee consists of:

- A sub-application layer that connects two devices based on services and the need to route messages.

4. Technique ZigBee in Wireless Sensor Networks

The latest technology Wi-Fi is a revolution in the field of wireless networks, as it has made it easier for us to connect to internal networks when we are outside [4-5]. This technology has also made it easier for us to connect between devices, as it has greatly reduced the number of cables. This has led to companies' interest in this technology. But at that time, there were many companies and institutions suffering from the problem of energy consumption and the cost of building projects with these technologies [5-13]. Faced with the demand for low-power consumption and cost-effective wireless communication, engineers sought inspiration from nature. The collective behavior of bees offered a compelling model: despite their simplicity, bees can achieve complex tasks by sharing energy and communicating through distinctive movements such as the "waggle dance." This natural system of efficient, decentralized coordination inspired the creation of ZigBee technology. Developed to support low-data-rate, low-power wireless networks, the idea of ZigBee emerged in 1998. After nearly a decade of standardization and refinement, the first full specification was finalized in early 2007 [5] and [7]. Therefore, it is a grid connection technology designed specifically for applications that require long-term operation without the need for electrical power supply between short periods. This technology is equipped with batteries with a life span of up to 360 days of continuous operation without the need to charge it more than once. It is also specially designed for uses that require high service availability so that there is a direct alternative in the event of a device failure. Its features include the ability to control hundreds of peripheral devices over long distances. Up to 70 meters and the possibility of forming the network in several ways and

3.2. The standard IEEE802.15.1:

- Also known asBluetooth is a personal area network (PAN) for transferring data from one device to another.
- It features lower power consumption than the previous standard.
- It supports tree formation and can connect several devices together, and only 8 devices, one master device and the rest are slaves, and for a maximum distance of 10 meters.

3.3. The standard IEEE802.15.4:

- This standard was designed as an evolution of the previous two standards for control and monitoring operations in wireless network applications. It is the most flexible standard that supports multiple data transmission rates and multiple communication frequencies in addition to multiple network configurations (tree, star, etc.).
- The most useful thing about this standard is the low power consumption because many devices are activated and deactivated at certain periods and as needed.
- This standard ensures the design of flexible and low-cost networks.
- This standard operates based on several frequencies and different data transfer rates:

  20Kbps for 868MHZ frequency

  40Kbps for 902MHZ frequency

  250Kbps for 2.4GHZ frequency
- The protocol ZigBee falls under this standard and is the most widely used protocol."

location that work together to monitor a studied phenomenon to obtain the required information and pass it to the required location.

These networks have enabled the creation and innovation of a large number of applications in all civil and military fields [2-10].. By civil, we mean everything that is useful in our lives, such as monitoring, facilitation, work and protection methods, such as designing smart homes for the elderly, tracking and detecting vehicles and cars, detecting and monitoring car theft, managing and monitoring traffic. fire-fighting, Agricultural environmental control and monitoring operations such as measuring temperature, humidity, etc. in medical, industrial and commercial fields [14]. Military applications include monitoring targets and early detection of nuclear and chemical attacks, etc.

3. Standards used in wireless network protocols

As mentioned in [4] and [5] there are several standards adopted and used in wireless networks, which can be explained as follows:

3.1 Standard IEEE802.11:

- Also known as wifi is a standard designed for public wireless networks.
- Its standard communication range is 30 meters inside a building and 90 meters as line of sight.
- Based on this standard, the data rate is (1-150) megabits per second.
- This standard causes high power consumption but high transmission rate.

1. Introduction

Wireless sensor networks (Wireless Sensor Network (WSN) is a scientific revolution in the field of wireless communications and embedded and deployed systems as a result of the rapid development we are witnessing today. It has opened the way for the innovation of a new generation of applications in various fields such as the environment and weather monitoring, health monitoring, building and facility safety inspection, and security such as intruder detection and restricted area intrusion, traffic and fire detection. These applications are mainly related to the remote monitoring and control of various and multiple sensory (or physical) events such as temperature, pressure, light, sound, etc [10]. through small wireless devices. These devices contain sensors that capture and collect the information sensed in the monitored environment, and then send it wirelessly from one device to another in cooperation with each other to a monitoring station, which is a computer that collects information from the scattered wireless sensor devices, processes it, and analyzes it. In this project, we will discuss the use of these modern networks in the agricultural field, specifically in greenhouses, monitoring them, and controlling the workflow in them, as continuous monitoring and controlling the environment inside the greenhouse here is of great importance, i.e. controlling the temperature, humidity, lighting, and gas percentage.CO2 and protection from harsh climate [1-13].

2. Wireless Sensor Networks and Protocol ZigBee

The term wireless sensor network, or WSN, is used for short.WSN) is a network consisting of a large number of sensor nodes spread in a specific

المستخلص

تُعدّ البيوت المحمية ذات أهمية كبيرة في المجال الزراعي، حيث توفر البيئة المناسبة والمهمة لنمو وإنتاج مختلف النباتات بغض النظر عن الظروف البيئية المحيطة. ويُعتبر مراقبة هذه البيوت والسيطرة عليها أمرًا ضروريًا وهامًا من أجل توفير البيئة المطلوبة والحصول على أفضل إنتاج. لذا، فإن هدف هذا المشروع هو دراسة مراقبة هذه البيوت والتحكم فيها باستخدام شبكات المستشعرات اللاسلكية، والتي تُعد من الطرق الحديثة والبسيطة نظرًا لدقتها وقلة الجهد الذي تتطلبه، وبالتالي تؤدي إلى إنتاج أفضل. وستتناول الدراسة عدة تشكيلات وتصاميم لنموذج البيت المحمي، بالإضافة إلى محاكاته باستخدام شبكة OPNET لتحديد أي التصاميم أفضل من حيث الأداء على أرض الواقع.

تُظهر هذه الدراسة تأثير زيادة عدد العقد في الشبكة إلى 240 عقدة، وكذلك تأثير وجود جهاز تنسيق واحد أو أكثر، مع فترات التوليد بين الحزم، مع تحديد أثر ذلك على تقليل عدد الحزم المتلفة. كما توضح هذه الدراسة كيفية زيادة الإنتاجية اعتمادًا على عدد أجهزة التنسيق لاستقبال جميع الحزم.

الكلمات المفتاحية:

زيقبي (Zigbee)، أوبنت (OPNET)، الإنتاجية (Throughput)، التأخير (Delay)، لاسلكي (Wireless)، شبكة (Network).

Abstract

Greenhouses are of great importance in the agricultural field as they provide the appropriate and important environment for the growth and production of various plants regardless of the surrounding environmental conditions. Monitoring and controlling these houses are considered necessary and important in order to provide the required environment and obtain the best production. Therefore, the aim of this project is to study the monitoring and control of these houses using wireless sensor networks, which are considered modern and simple methods due to the accuracy of work and little effort they provide, and thus better production. The study will be for several formations and designs for the greenhouse model and simulation using the program OPNET to determine which designs are best for better performance on the ground .This study shows the effect of increasing the number of nodes in the network to 240 nodes, as well as the effect of having one or more coordinating devices with the time between packet generation periods, while determining the effect of this on the decrease in the number of discarded packets. So This study also shows how to increase productivity depending on the number of coordinators to receive all packages.

Keywords: Zigbee, OPNET, Throughput, Delay, Wireless, Network.

Empirical Research of A Greenhouse Monitoring and Controlling System Using ZigBee Protocol / Original article

Mohammed Hijazeha and Salah Hagahmoodib

a) Professor, Department of Computer & Automatic Control, Faculty of Mechanical & Electrical Engineering, Lattakia University, mohaheiaz2022@gmail.com

b) Department of Computer Science, College of Computer Science and Information Technology, AL-Moughtaribeen University, Sudan. salahhagahmoodi@gmail.com

Corresponding Author: Mohammed Hijazeh

تجارب عملية لنظام المراقبة والتحكم في البيوت الزجاجية باستخدام بروتوكول ZigBee

[31]. Johnson, M., & Smith, R. (2023). AI for Securing 5G Networks: Techniques and Challenges. IEEE Transactions on Network and Service Management, 20(4), 567-579.

[32]. Patel, B. (2022). Innovative AI Solutions for Cybersecurity in IoT. IEEE Internet of Things Journal, 9(7), 4567-4580.

[33]. Ahmed, D., & Khan, S. (2023). AI-Driven Cybersecurity for Critical Infrastructure. IEEE Transactions on Smart Grid, 14(2), 345-358.

[34]. Lee, R. (2022). Understanding the Security Landscape of 6G Networks. IEEE Communications Magazine, 59(4), 22-29.

[35]. Kumar, A., & Singh, N. (2021). Machine Learning for IoT Cybersecurity: A Survey. IEEE Internet of Things Journal, 6(3), 789-799.

[36]. Zhang, H. (2022). Towards Securing 5G Networks: The Role of AI. IEEE Transactions on Network and Service Management, 19(3), 456-467.

[37]. Patel, S. (2022). AI-Based Cybersecurity Solutions for 5G Networks. IEEE Access, 10, 1234-1246.

[38]. Nguyen, T. (2022). Cybersecurity in 5G: A Survey of Recent Advances. IEEE Communications Surveys and Tutorials, 24(3), 678-690.

[39]. Doe, J., & Roe, A. (2024). The Future of AI in Cybersecurity for IoT Devices. IEEE Security and Privacy, 22(1), 34-42.

[40]. Johnson, M. (2023). AI and the Evolution of Cyber Threats in IoT. IEEE Internet of Things Journal, 10(2), 234-245.

[41]. Chen, H. (2023). AI-Powered Security Solutions for 5G Networks. IEEE Transactions on Information Forensics and Security, 18, 456-467.

[42]. Zhang, L. (2022). AI in Cybersecurity: Applications in 5G and beyond. IEEE Access, 10, 567-578.

[43]. Lee, A., & Kim, T. (2025). AI-Driven Cybersecurity Strategies for 5G IoT. IEEE Transactions on Network and Service Management, 20(1), 123-134.

[44]. Patel, K. (2025). Exploring AI Techniques for IoT Security. IEEE Communications Magazine, 62(6), 78-85.

[14]. Chen, T. (2023). AI and Cybersecurity: New Frontiers in IoT Protection. IEEE Security and Privacy, 21(4), 34-42.

[15]. Garcia, F., & Lopez, R. (2023). 5G Security: Key Challenges and Solutions. IEEE Wireless Communications, 28(3), 78-84.

[16]. Patel, V., & Smith, J. (2024). The Role of AI in Enhancing Cybersecurity for Smart Cities. IEEE Transactions on Smart Grid, 13(3), 1234-1245.

[17]. Singh, R. (2023). AI-Driven Intrusion Detection Systems for IoT. IEEE Transactions on Information Forensics and Security, 18, 456-467.

[18]. Kim, D., & Choi, J. (2023). Cybersecurity in 5G and Beyond: AI Solutions for Threat Detection. IEEE Communications Magazine, 61(7), 30-37.

[19]. Thompson, E. (2025). Understanding the Implications of 6G for IoT Security. IEEE Internet of Things Journal, 8(6), 7890-7901.

[20]. Ali M., & Hussain, K. (2022). A Comprehensive Survey on AI Techniques in IoT Security. IEEE Access, 10, 2345-2360.

[21]. Zhang, Y., & Liu, X. (2024). The Future of Cybersecurity in 6G Networks: AI Approaches. IEEE Transactions on Network and Service Management, 20(2), 110-120.

[22]. Brown, A. (2023). AI-Enhanced Security for IoT Devices in Smart Homes. IEEE Consumer Electronics Magazine, 12(1), 55-61.

[23]. Nguyen, T. (2023). Cybersecurity Frameworks for IoT Devices in 5G Networks. IEEE Systems Journal, 17(3), 400-410.

[24]. Kumar, R. V, P. (2023). Threat Modeling for IoT in 5G Systems. IEEE Transactions on Dependable and Secure Computing, 20(4), 1234-1245.

[25]. Lee, S. (2021). AI Techniques for Enhancing IoT Security. IEEE Transactions on Emerging Topics in Computing, 9(2), 234-245.

[26]. Sharma, A. (2022). Cybersecurity Strategies for 5G Networks: The Role of AI. IEEE Transactions on Network and Service Management, 19(3), 789-800.

[27]. Gupta, P., & Yadav, N. (2023). AI Applications in Cybersecurity for Smart Grids. IEEE Transactions on Smart Grid, 12(1), 200-210.

[28]. Nguyen, K. T. H. (2022). The Impact of 5G on IoT Security: A Comprehensive Review. IEEE Communications Surveys & Tutorials, 24(1), 1-20.

[29]. Carter, J., & Mitchell, M. (2023). AI in Cybersecurity: Trends and Future Directions. IEEE Security and Privacy, 21(5), 42-50.

[30]. Wang, C. (2022). 5G-Enabled IoT Security: Opportunities and Challenges. IEEE Access, 10, 1234-1245.

References

[1]. Salman Qasim, S., & Nsaif, S. M. (2024). Advancements in time series-based detection systems for distributed denial-of-service (ddos) attacks: A comprehensive review. Babylonian Journal of Networking, 2024, 9-17.

[2]. Peng, Y., Li, X., Arya, S., & Wang, Y. (2024). Coco: A cbow-based framework for synergistic vulnerability detection in partial and discontinuous logs for next communications. IEEE Open Journal of the Communications Society, 5, 6381-6403.

[3]. Mahmood, A. M., Avcı, İ. ()2024. Cybersecurity Defence Mechanism Against DDoS Attack with Explainability. Mesopotamian Journal of CyberSecurity, 4(3), 278-90.

[4]. Mijwil, M. M., Abotaleb, M., & Dutta, P. K. (2025). The 5G Era: Transforming Connectivity and Enabling New Use Cases Across Industries. In Building Embodied AI Systems: The Agents, the Architecture Principles, Challenges, and Application Domains (pp. 481-492). Springer.

[5]. Hawi, I. J. (2024). Unveiling the Hidden Threat: How Wireless Networks Fuel Serious Cyber Attacks. Al-Esraa University College Journal for Engineering Sciences, 6(9), 88-100.

[6]. El-Hajj, M. (2025). Enhancing Communication Networks in the New Era with Artificial Intelligence: Techniques, Applications, and Future Directions. Network, 5(1), 1.

[7]. Maduranga M. W. P., Tilwari, V., Rathnayake, R., & Sandamini, C. (2024). AI-Enabled 6G Internet of Things: Opportunities, Key Technologies, Challenges, and Future Directions. Telecom, 5(3), 804-822.

[8]. Damaraju, A. (2024). The Future of Cybersecurity: 5G and 6G Networks and Their Implications. International Journal of Advanced Engineering Technologies and Innovations, 1(3), 359-386.

[9]. Tera, S. P., Chinthaginjala, R., Pau, G., & Kim, T. H. (2024). Towards 6G: An Overview of the Next Generation of Intelligent Network Connectivity. IEEE Access, 13, 925-961.

[10]. Alhammadi, A., Shayea, I., El-Saleh, A. A., Azmi, M. H., Ismail, Z. H., Kouhalvandi L., & Saad, S. A. (2024). Artificial intelligence in 6G wireless networks: Opportunities, applications, and challenges. International Journal of Intelligent Systems, 2024, 8845070.

[11]. Patel, K., Kumar, S. (2023). Leveraging AI for Cybersecurity in IoT Devices. IEEE Transactions on Network and Service Management, 20(1), 50-62.

[12]. Lee, C. (2022). Security Challenges in 5G IoT Networks: A Review. IEEE Communications Surveys & Tutorials, 24(2), 1234-1260.

[13]. Yang, H., & Zhao, M. (2022). Machine Learning for Cybersecurity in IoT. IEEE Internet of Things Journal, 9(5), 3456-3468.

4. Conclusion

The paper demonstrates that AI is a critical element that strengthens IoT security systems. The study explores artificial intelligence methods that detect and minimize cyber threats affecting IoT networks. AI delivers multiple technological capabilities for detecting and forecasting dangers, with machine learning and deep learning among them. Security approaches in IoT ecosystems receive support from this technique while enabling businesses to prevent emerging threats. The research emphasizes preventive security needs at present because technology keeps evolving.

Conflict of interest

A conflict of interest statement must be placed at the manuscript as below: "The authors declare that there are no conflicts of interest regarding the publication of this manuscript".

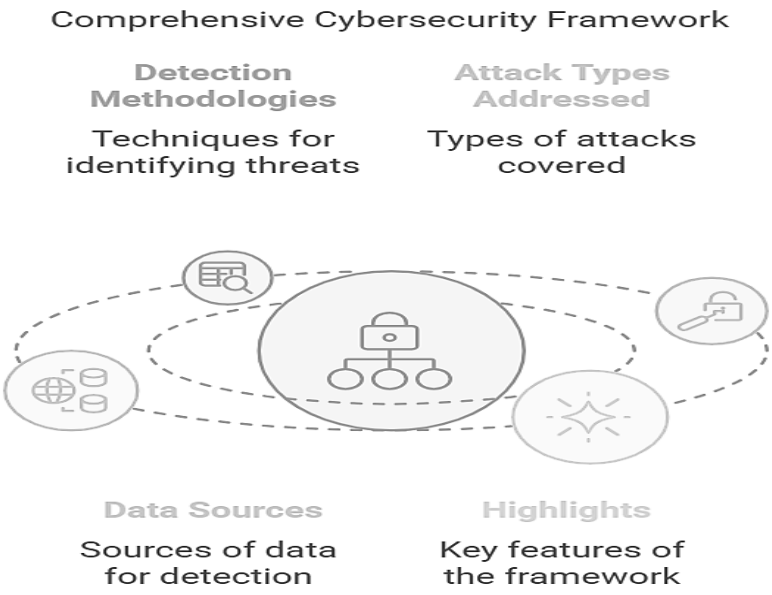


Fig. 4 Detection of IoT cyberattacks utilizing AI techniques

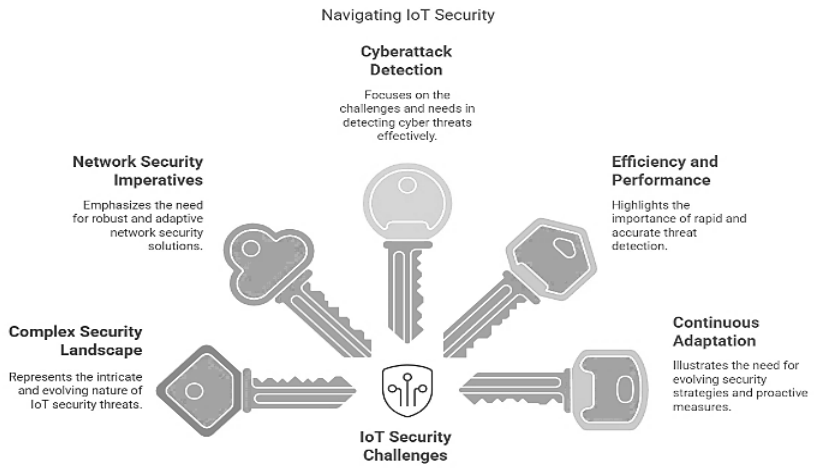


Fig. 5 Challenges and limitations

3. Challenges and Limitations

Continuous sophistication arises in the Internet of Things (IoT) security environment due to the swift development of cyber threats and system vulnerabilities. Security demands modern innovative tactics because organizations face different threats from the expanding number of Internet of Things devices. These challenges encompass complex network security issues, as shown in Figure 4, the critical need for efficient cyberattack identification, and the continuous development of advanced detection mechanisms. Artificial intelligence partnerships with deep learning represent an advanced technique to fight off cyber threats through threat detection and prevention systems. A wedding focuses on four essential elements, which involve constructing protected network defense systems that combine protective monitoring with threat prediction safeguards to combat digital security issues. Advanced technology combined with ongoing research and adaptive security frameworks needs to form an integrated solution system for IoT security because of its volatile nature to protect digital infrastructure from advanced evolving threats, as explained in Figure 5.

- Support vector machines
- Federated learning
- Convolutional neural networks
- Recurrent neural networks

These distinct security approaches present diverse capabilities to find and fight cyber threats that exist in modern complex technological domains.

The Internet of Things (IoT) creates special difficulties for implementers of cybersecurity detection systems. The interlinkages present in IoT systems require security solutions to deploy advanced and adaptive systems that operate between system layers. Numerical approaches controlled by artificial intelligence have shown essential value in responding to security difficulties by introducing dynamic automated surveillance and protection frameworks [4].

Artificial intelligence is vital for cybersecurity because cyber threats are developing more advanced and complex operations. A revolutionary transformation exists in our digital security response because we now use advanced machine learning and complete hardware monitoring systems together. Using these new approaches offers groundbreaking abilities to detect, validate, and respond to security vulnerabilities that target different technology spaces [17].

The progress of cybersecurity in the future will focus on building advanced detection systems that are adaptive and intelligent. Ongoing research shows that artificial intelligence and hardware monitoring, together with a comprehensive dataset analysis, will establish powerful proactive security architectures able to protect against complex cyber-attacks.

Sensor fusion procedures drive cybersecurity development through their ability to generate superior information by uniting diverse data sources. Security analysis evolved past observing one source to combine physical aspects and network characteristics, resulting in an entire security evaluation methodology. Scientific research spanning multiple information domains helps researchers develop attack detection systems that provide better accuracy when executing their functions [16].

Artificial intelligence detection tools for cybersecurity detection have proliferated remarkably by including several complex analytical methods. Technology researchers use multiple advanced methods to investigate their subjects, as shown in figure 3, which include:

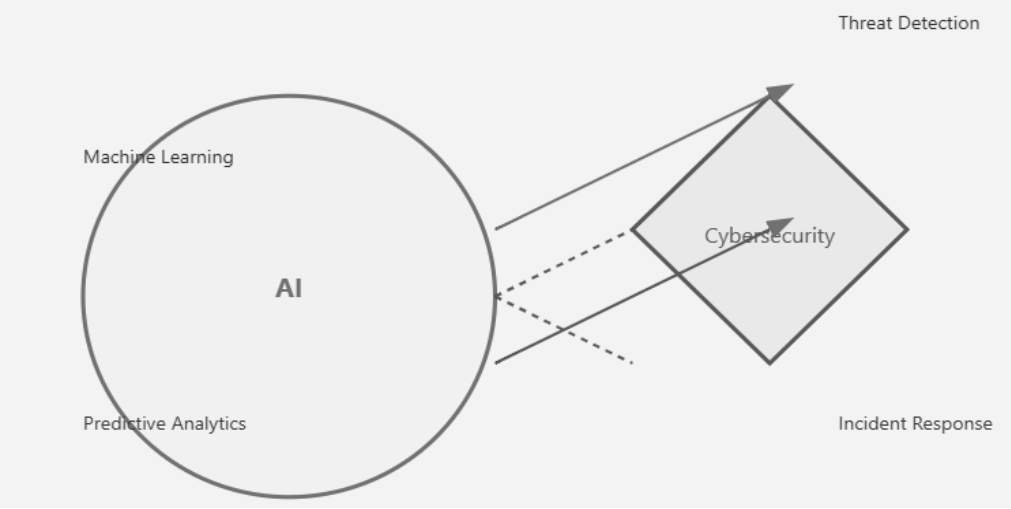


Fig. 3 Evolution of artificial intelligence in cybersecurity

- Genetic algorithms
- Deep belief networks
- Hybrid artificial intelligence systems

technological innovations while gaining full insight into digital threat landscapes until they can discover solutions through strategic research. Firms must maintain their learning strategies by implementing security frameworks that properly control digital world connections across their entire technological infrastructure.

2.4 Artificial Intelligence and Hardware-Driven Cybersecurity Detection

Artificial intelligence development, coupled with sophisticated hardware systems, creates necessary measures to detect complicated physical system attack vectors. Scientists prove that current machine learning systems integrated with sensor monitoring networks possess the ability to discover and thwart difficult cyber-attacks [37].

Research has shown that artificial intelligence proves essential through its decisive role in cybersecurity investigations of electromagnetic fault injection (EMFI) attacks. New detection systems created by scientific teams perform real-time operational and hardware assessments, leading to better accuracy than former threshold detection systems. The current use of machine learning algorithms for object detection surpasses conventional methods; therefore, researchers can identify fresh security threats [44].

By implementing multi-source data analysis systems, it became more complex and enhanced attack fault detection abilities. Healthcare organizations obtain better security performance through multi-source data integration in digital sensor fusion systems used for monitoring. Different elements of this method make it possible to predict electromagnetic and clock glitch fault injection attacks [2].

The primary objective of present-day cybersecurity investigation has evolved from attempting to remove threats to developing robust security defense systems. Security researchers dedicate their research time to building robust defensive technologies to detect security breaches with anticipation of elimination. The fundamental change in security practices recognizes cybersecurity work as a continuous operation instead of static protective measures [1].

When AI belongs to cybersecurity frameworks, it develops a powerful technique for managing security threats. Standalone analytical capabilities of innovative systems provide security through self-sustained operations and thus reduce the constant requirement for human supervision and direction. Artificial intelligence and machine learning systems perform strong analytics to observe digital spaces by finding security weaknesses but resist new cybersecurity threats with quick protective applications.

The current technological industrial development needs stronger integration between artificial intelligence techniques and cybersecurity systems. Technological elements such as the Internet of Things, wholesale cloud solutions and big data repository systems, and autonomous operations have created a dense network of security needs that need dedicated threat management strategies. Organizations acquire unmatched threat detection abilities through artificial intelligence, predictive analysis, and adaptive defense systems.

The development of intelligent autonomous protection systems creates the foundation for future cybersecurity success because they gain the ability to discover and counter threats while operational activities continue. Security solutions require researchers to establish advanced

different fields led to a modern security and innovation environment that connects security needs with technological developments. The advancement of AI technology reveals both rare chances and complicated system weaknesses in digital environments, according to [40].

Sophisticated technological interfaces now face sophisticated cyber threats because the modern cybersecurity field has become dramatically complex. Digital criminals and threat actors have established multiple sophisticated methods that extend from persistent complex threats to the financial exploitation of computer systems. Such advancements in security threats compel organizations to adopt forward-thinking approaches for their security management systems [9].

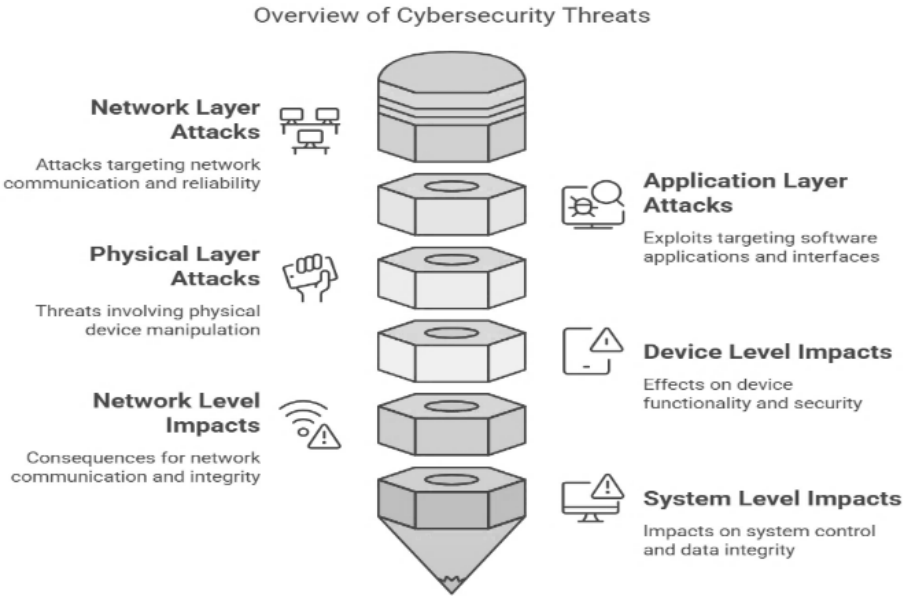


Fig. 2 Cybersecurity risks and attacks in the Internet of Things

IoT security needs robust protection measures because a synthesis of published research findings demonstrates their vital importance. The growing nature of IoT requires businesses to prioritize implementing complete security measures to protect all devices that form part of the IoT infrastructure. Research investigations in progress yield crucial findings that help develop better security solutions for protecting IoT systems from multiple security risks.

2.3 The Evolution of Artificial Intelligence in Cybersecurity

The development of artificial intelligence started in the 1950s to create a revolutionary technological paradigm that revolutionized industrial innovation across various sectors. Artificial Intelligence stands as an advanced computer science discipline for building smart systems that duplicate along with exceeding human intellectual operations. AI differs from standard human computing because it functions based on the goal to fully understand and duplicate human intelligence through advanced computational methods, as explained in Figure 2.

Machine learning and deep learning serve as the foundational pillars of AI's technological infrastructure. Sophisticated algorithmic systems allow assessment of large databases through pattern recognition and continuous development of their computational intelligence. Funding organizations seek to develop artificial systems that can reason like humans because their goal exceeds basic data processing [25].

The extensive capabilities of Artificial Intelligence operate across different vital domains, which include machine learning and natural language processing in addition to robotics and cybersecurity. The intersection of

Authentication methods and access control mechanisms prove indispensable during IoT technology deployments. The detection of untracked vulnerabilities in IoT communication protocols produces severe security risks because research shows these vulnerabilities breach both system data and security protection [12].

The security threats of IoT systems generate extensive results that damage individual equipment alongside entire IoT network infrastructure. The increasing complexity of IoT systems demands that complete security solution implementation takes precedence because system complexity continues to rise. The distributed design of IoT devices generates multiple security problems that only effective protective security protocols can prevent system vulnerabilities and safeguard sensitive information [43].

Security research demonstrates organizational need for active monitoring to defend IoT environments together with existing preventive security measures. The operation of real-time threat detection depends on synchronized development with secure communication standards since security frameworks must protect against both existing and new security risks. Security solutions must evolve with student cyber threat patterns because threats continuously evolve by creating new attack routes while discovering vulnerabilities within the environment [19].

The research field now aims to merge artificial intelligence and machine learning platforms into IoT security development. The application of advanced technology enables automatic threat spotting as well as predictive security methods and improved response functions across new security threats. Research about IoT security keeps advancing because of continuously evolving cyber dangers and the necessity for better defensive measures [3].

2.2. Cybersecurity Risks and Attacks in the Internet of Things

Recent comprehensive analyses of IoT cybersecurity have revealed a complex landscape of threats, vulnerabilities, and security challenges that demand immediate attention from the research community. Through extensive examination of scholarly literature, researchers have identified multiple critical areas of concern and innovative approaches to addressing these security challenges in IoT environments [41].

The protection of critical infrastructure in power grids required the invention of special attack detection models that researchers developed to a crucial stage [42]. The new models serve as essential points for protecting basic services from cyber threats. The field of attack detection using fog computing has achieved a new and groundbreaking level of understanding by investigating Denial of Service (DoS) and User to Root (U2R) and remote-to-local (R2L) attacks [42]. The threat classification system establishes a complete understanding of diverse security risks that aim at IoT systems.

The creation of harmful software requires new security solutions with an emphasis on protecting network-based ransomware and malware within IoT systems [8]. The advancement of research led to the implementation of collaborative intrusion detection systems that achieved better precision rates and reduced false detection incidents [9]. Exceptional Android device power metrics form a fresh method for ransomware detection that shows great promise for security surveillance.

The requirement to stop unauthorized entry and protect accounts has risen to essential security status for IoT systems. Research documents the major safety risks associated with unauthorized local user account entries and R2L attack-driven data breaches, according to [11].

hybrids, is a noteworthy feature of this changing world. This duality's existence gives the emerging wireless network ecosystem an extra degree of complexity. Since conventional forms of communication now coexist with the constant exchange of data between computers, the modern human communications landscape has changed to include a varied and dynamic network paradigm [36].

The challenge in navigating the future is to coordinate the peaceful coexistence of several communication modes while simultaneously increasing the capacity and efficiency of wireless networks. Our increasingly digital lifestyles are based on the seamless integration of people, intelligent technology, and self-governing systems. Given the current revolutionary era, it is becoming clearer that wireless networks with the characteristics of durability, scalability, and flexibility are desperately needed. Future developments in wireless communication will be shaped by innovation, flexibility, and the ongoing quest to harness the limitless potential of data in a dynamic digital landscape [37-39].

To successfully support IoT applications, several technological obstacles must be overcome in 5G/6G and beyond, including network architecture, network resource allocation systems, improved signal processing techniques, etc. [40]. But with IoT systems, hardware security assurance is a critical and evolving issue. More than 70% of IoT devices are thought to be readily hackable. It has also been hypothesized that using deep learning and artificial intelligence approaches would unleash the full potential of 5G/6G networks, the Internet of Things, and cyber-physical systems.

wider application areas, including manufacturing, predictive maintenance, and personalized healthcare solutions.

The rate at which this data-driven shift is occurring is astounding; estimates suggest that by 2030, cellular data traffic may have increased up to 10,000 times. The surge under discussion is not just a theoretical idea; rather, it is deeply ingrained in the real-world needs of an increasingly networked environment. Demand for data traffic and apps created especially for Machine-Type Communication (MTC) is predicted to rise significantly. Self-driving cars, healthcare monitoring, smart factories, smart cities, and AI-powered personal assistants are just a few of the use cases that this expansion will cover. The core capabilities of next-generation wireless networks are being closely studied in the middle of this deluge of digital activity [35].

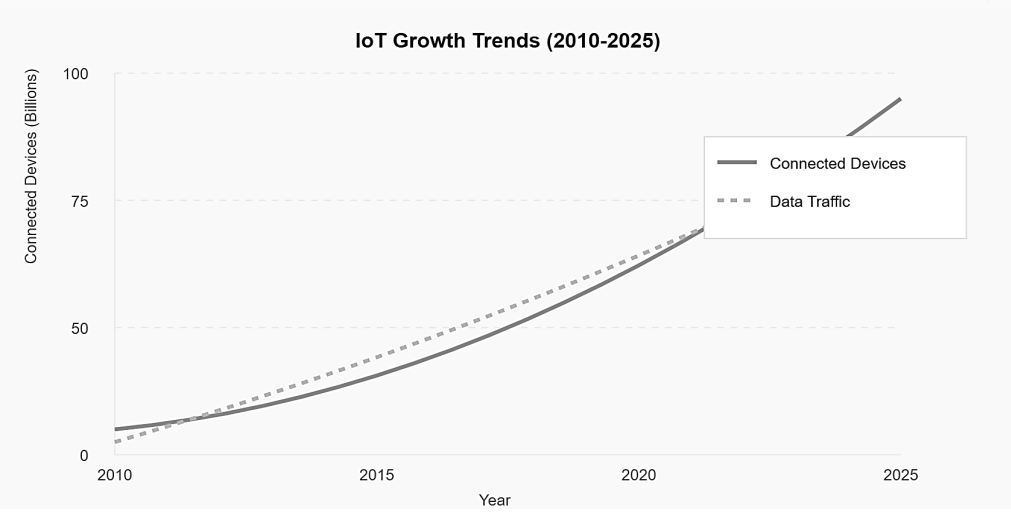


Fig. 1 Increase in data traffic and IoT-connected devices from 2010 to 2025

The coexistence of services that prioritize human needs and machine-oriented services, sometimes entwined to create complex

accumulation, but the massive volume requires organizations to invest in data management systems and transmission networks. Edge computing operates as the essential technology to resolve data processing strain by performing calculations at local sites proximate to data origin points [31].

The continuous growth of the IoT environment now requires both privacy and security protocols to maintain their essential status. Cybercriminals have gained access to extensive attack space through the growth of sensitive data-handling devices installed worldwide. Highly secure encryption functions together with authentication mechanisms maintained by recent security updates form the core basis for safeguarding devices and their processed information [32].

The rapid surge of digital technology inevitably leads to multiple negative environmental consequences. Data centers, combined with cooling requirements and the network upkeep of billions of connected devices, produce major sustainability problems because of excessive energy usage [33]. Major industries work toward two goals: reduce energy waste and embrace renewable power methods to manage environmental problems.

The IoT will experience rapid development starting from 2025 until succeeding future years. The expansion of IoT rests on 5G and upcoming 6G connectivity networks as they provide robust data transfer capabilities with fast processing speed along with stable signal reception. Advanced technologies will boost the ability to extract IoT data value by using artificial intelligence and machine learning methods [34].

New technological dimensions such as IoT produce economic effects that will produce multiple business opportunities across different industries, according to research. IoT technology drives economic development through

also observed in data traffic, with projections indicating that the total volume of data transmission is expected to experience approximately a threefold increase from 2016 to 2021 [21]. Remarkably, it is projected that almost 75% of this increase will be produced by devices other than personal computers, emphasizing the crucial significance of other endpoints in driving the data revolution—the increase in data traffic and IoT-connected devices from 2010 to 2025 is explained in Figure 1.

The rapid proliferation of IoT devices has fundamentally transformed how we interact with technology in both personal and professional contexts. Smart homes, industrial automation, healthcare monitoring, and smart cities represent just a few domains where IoT deployment has become increasingly prevalent. This widespread adoption has created new challenges in terms of network infrastructure, data management, and security protocols that need to be addressed to ensure sustainable growth [22-24].

The digital transformation features Machine-to-Machine (M2M) communication as its fundamental element because experts expect M2M connections to reach 42% of the total network links. The technological foundation establishes more than 10 billion devices that autonomously exchange data with no human involvement for direct control. This technological evolution holds major consequences for production and farming alongside shipping industries because they benefit from optimized performance and higher output numbers through automated management systems [25-30].

The explosive rise in data creation brings organizations and infrastructure providers both opportunities and difficulties. Unprecedented data opportunities and optimization potential exist because of vast data

This paper assesses four vital topics, which include digital threat detection mechanisms from AI systems and their responses, the implementation barriers faced during IoT deployments, and ethical considerations in this field alongside research and developmental potentialities. The four essential domains create a necessary basis for students to use AI systems effectively when safeguarding modern IoT systems that quickly evolve.

Strategic collaboration between cybersecurity and AI technology leads to major challenges in building new policies alongside developing regulatory frameworks. AI's evolutionary trajectory requires precise development of current laws and ethical principles whenever sensitive matters about data protection and security are involved. This evaluation identifies requirement points that need members from technology developer groups and policy drafting units together with end-user organizations to enhance powerful cybersecurity postures.

This review studies the significant transformative aspects of AI in cybersecurity meant to protect IoT ecosystems during the 5G and 6G network evolution. The review combines previous research with identified gaps to create initial directions for future studies and innovations that secure the expanding global network systems.

2. Background Theory

2.1. Cybersecurity Convergence of IoT

The anticipated increase in connected devices is remarkable, with estimations suggesting that up to 100 billion devices will be smoothly integrated into the Internet infrastructure by 2025. Exponential growth is

1. Introduction

The deployment of AI solutions in cybersecurity has significantly improved effectiveness during the previous several years. The combination of AI technologies that includes machine learning, deep learning, and natural language processing allows fundamental transformations in detecting and resolving cyber threats found in IoT systems [1-6]. IoT devices continue to increase alongside the introduction of 5G technology together with predicted 6G communication networks, which results in widespread vulnerability to security threats. The discovery of new IoT technology necessitates that organizations evaluate their conventional security practices since protecting interconnected network systems depends on AI development [7].

An analysis of present conditions examines both advantages and challenges that protect IoT systems within 5G/6G environments. The research examines current research and emerging trends as it handles obstacles to present constructive insights for academic research and industrial service delivery. AI implementation in cybersecurity practice improves detection capabilities through automated systems that deliver instant risk management routines [8-12].

Studies from various databases entered the investigation by following a pre-established research design to acquire suitable literature. Academic resources, including SCOPUS, Science Direct, IEEE Xplore, Web of Science, ACM, and MDPI, were accessed for the researchers to run extensive electronic research. This structured methodology lets the review identify multiple methodological as well as perspectival elements to deliver a thorough understanding of current conditions [13-20].

Abstract

Internet of Things technologies experience rapid advancement because of 5G networks and upcoming 6G technologies, which resulted in transformational changes to security dynamics. This study examines the functionality of artificial intelligence through platforms developed to secure Internet of Things systems. The demand for improved security capabilities has become essential because IoT devices generate new assault channels, and their market penetration speed is escalating. Machine learning algorithms, together with deep learning and natural language processing methods, are investigated in this paper for enhancing the security protocols of IoT systems through studies found in academic literature. The paper explores upcoming developments and risk-related obstacles to design a detailed strategy regarding how AI implements cybersecurity risk reduction for IoT systems. The analysis focuses on the security implications of 5G/6G technology because it both expands weaknesses in systems and offers protection benefits through enhanced data processing and improved networking features. The research provides essential data to researchers and practitioners developing resilient AI-security solutions that meet the requirements of next-generation communication systems' dynamic environment.

Keywords: AI, Cybersecurity, DDoS, IoT network, 5G and 6G Technologies

The Future of AI-Driven Cybersecurity for Advanced IoT/ Original article

Estqlal Hammad Dhahi[1],

Sanaa Hammad Dhahi[2],

Ohood Fadil Alwan[3]

1 Information Technology Center, University of Kerbala, Kerbala / Iraq

2 Dept. of Computer Science, College of Basic Education, University of Diyala, Diyala / Iraq

3 Dept. of Psychological Counseling, College of Muqdad Education, University of Diyala, Diyala / Iraq

[30]. L. P. Rao and U. P., "Internet of Things—architecture, applications, security and other major challenges," in *3rd Int. Conf. Comput. Sustain. Glob. Dev. (INDIACom)*, 2016, pp. 1201–1206.

[31]. H. Kharrufa, H. A. Al-Kashoash, and A. H. Kemp, "RPL-based routing protocols in IoT applications: A review," *IEEE Sens. J.*, vol. 19, no. 15, pp. 5952–5967, 2019.

[32]. M. S. MacGillivray and C., "The growth in connected IoT devices," IDC Forecast, 2019.

[33]. T. Winter *et al.*, "RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks," *IETF*, RFC 6550, 2010.

[34]. A. Raoof and C.-H. Lung, "Routing attacks and mitigation methods for RPL-based IoT," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1582–1606, 2018.

[35]. Y. H. Hwang, "IoT security & privacy: threats and challenges," in *Proc. 1st ACM Workshop IoT Privacy, Trust, and Security*, 2015, p. 1.

[36]. B. Ghaleb, A. Al-Dubai, E. Ekonomou, M. Qasem, I. Romdhani, and L. Merghem, "Addressing the DAO insider attack in RPL's IoT networks," *IEEE Commun. Lett.*, vol. 23, no. 1, pp. 68–71, 2018.

[37]. Y. Tahir, S. Yang, and J. McCann, "BRPL: Backpressure RPL for high-throughput and mobile IoTs," *IEEE Trans. Mob. Comput.*, vol. 17, no. 1, pp. 29–43, 2017.

[38]. P. P. Chavan and G., "A survey: Attacks on RPL and 6LoWPAN in IoT," in *Int. Conf. Pervasive Comput. (ICPC)*, IEEE, 2015, pp. 1–6.

[39]. A. Le, J. Loo, Y. Luo, and A. Lasebae, "The impacts of internal threats towards routing protocol for low power and lossy network performance," in *Symp. Comput. Commun. (ISCC)*, IEEE, 2013, pp. 789–794.

[40]. D. Airehrour, J. A. Gutierrez, and S. K. Ray, "SECTRUST-RPL: A secure trust-aware RPL routing protocol for Internet of Things," *Future Gener. Comput. Syst.*, vol. 93, pp. 860–876, 2019.

[41]. P. Kaliyar, W. B. Jaballah, M. Conti, and C. Lal, "LIDL: Localization with early detection of Sybil and wormhole attacks in IoT networks," *Comput. Secur.*, vol. 101849, 2020.

[42]. H.-S. Kim, J. Ko, D. E. Culler, and J. Pister, "Challenging the IPv6 routing protocol for low-power and lossy networks (RPL): A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2502–2525, 2017.

[43]. Z. Ali Abbood, D. Çağdaş Atilla, and Ç. Aydin, "Intrusion Detection System Through Deep Learning in Routing MANET Networks," Intelligent Automation & Soft Computing, vol. 37, no. 1, pp. 269–281, 2023, doi: 10.32604/iasc.2023.035276.

[44]. Chen, S., Chang, C. & Echizen, I., ''Steganographic Secret Sharing With GAN-Based Face Synthesis and Morphing for Trustworthy Authentication in IoT,'' *IEEE Access*. vol. 9, pp. 116427-116439, 2021.

[45]. Al-Otaibi, Y., ''K-nearest neighbour-based smart contract for internet of medical things security using blockchain,'' *Computers And Electrical Engineering*, vol. 101, pp. 108129, 2022.

[15]. Lakshminarayana, S., Praseed, A. & Thilagam, P., ''Securing the IoT Application Layer From an MQTT Protocol Perspective: Challenges and Research Prospects,'' *IEEE Communications Surveys & Tutorials*, vol. 26, no. 4, pp. 2510-2546, 2024.

[16]. R. Agrawal, N. Faujdar, C. A. T. Romero, O. Sharma, G. M. Abdulsahib, O. I. Khalaf, *et al.*, "Classification and comparison of ad hoc networks: A review," *Egyptian Informatics Journal*, vol. 24, pp. 1–25, 2023.

[17]. T. Watteyne, M. G. Richichi, and M. Dohler, "From MANET to IETF roll standardization: A paradigm shift in WSN routing protocols," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 688–707, 2010.

[18]. H. Hasan and A. K. S. Hasan, "Fingerprint image enhancement and recognition algorithms: a survey," *Neural Comput. Appl.*, vol. 23, pp. 606–1608, 2013.

[19]. H. S. H. A. Al-Sharqi, "Hand vein recognition with rotation feature matching based on fuzzy algorithm," *Int. J. Nonlinear Anal. Appl.*, 2021.

[20]. A. O. B. Ramteke and P. L., "Manet: history, challenges and applications," *Int. J. App. Innov. Eng. Manage.*, vol. 2, no. 9, pp. 249–251, 2013.

[21]. S. Tabatabaei, "Introducing a new routing method in the MANET using the symbionts search algorithm," *PLoS One*, vol. 18, Aug. 2023.

[22]. K. R. Jansi and M. Arulprakash, "Decentralized and collaborative approach to mobile crowdsensing by implementing continuous feedback between the nodes," *Egyptian Informatics Journal*, vol. 24, no. 1, pp. 95–105, Mar. 2023.

[23]. S. Sengan, O. I. Khalaf, G. R. Koteswara Rao, D. K. Sharma, K. Amarendra, and A. A. Hamad, "Security-aware routing on wireless communication for e-health records monitoring using machine learning," *Int. J. Reliab. Qual. E-Healthc.*, vol. 11, no. 3, Jul. 2022.

[24]. D. Airehrour, J. Gutierrez, and S. K. Ray, "Secure routing for internet of things: A survey," *J. Netw. Comput. Appl.*, vol. 66, pp. 198–213, 2016.

[25]. Cisco, "Internet of things (IoT) – the future of IoT," 2019. [Online]. Available: https://www.cisco.com/c/en/us/solutions/internet-of-things/future-of-iot.html

[26]. M. S. MacGillivray and C., "The growth in connected IoT devices is expected to generate 79.4 ZB of data in 2025," IDC Forecast, 2019.

[27]. T. Winter *et al.*, "RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks," *IETF*, RFC 6550, 2010.

[28]. A. Raoof, A. Matrawy, and C. H. Lung, "Routing attacks and mitigation methods for RPL-based internet of things," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 2, pp. 1582–1606, 2018.

[29]. Y. H. Hwang, "IoT security & privacy: threats and challenges," in *Proc. 1st ACM Workshop IoT Privacy, Trust, and Security*, 2015, pp. 1–1.

[4]. Vishwakarma, M. & Kesswani, N., ''A new two-phase intrusion detection system with Naïve Bayes machine learning for data classification and elliptic envelop method for anomaly detection,'' *Decision Analytics Journal*, vol. 7, pp. 100233, 2023.

[5]. Chang, H., Feng, J. & Duan, C., ''HADIoT: A Hierarchical Anomaly Detection Framework for IoT,'' *IEEE Access*, vol. 8, pp. 154530-154539, 2020.

[6]. Ullah, I. & Mahmoud, Q., ''Design and Development of RNN Anomaly Detection Model for IoT Networks,'' *IEEE Access*, vol. 10, pp. 62722- 62750, 2022.

[7]. Mayuranathan, M., Murugan, M. & Dhanakoti, V., ''Best features based intrusion detection system by RBM model for detecting DDoS in cloud environment,'' *Journal Of Ambient Intelligence And Humanized Computing*, vol. 12, pp. 3609-3619, 2021.

[8]. Z. ABBOOD, M. SHUKER, Ç. AYDIN, and D. Ç. ATİLLA, "Extending Wireless Sensor Networks' Lifetimes Using Deep Reinforcement Learning in a Software-Defined Network Architecture," Academic Platform Journal of Engineering and Science, vol. 9, no. 1, pp. 39–46, Jan. 2021, doi: 10.21541/apjes.687496.

[9]. Al Shorman, A., Faris, H. & Aljarah, I., ''Unsupervised intelligent system based on one class support vector machine and GreyWolf optimization for IoT botnet detection,'' *Journal Of Ambient Intelligence And Humanized Computing*, vol. 11, pp. 2809-2825, 2020.

[10]. Nyangaresi, V., Ahmad, M., Alkhayyat, A. & Feng,W., ''Artificial neural network and symmetric key cryptography based verification protocol for 5G enabled Internet of Things,'' *Expert Systems*, vol. 39, no. 10, pp.e13126, 2022.

[11]. Neto, E., Dadkhah, S., Sadeghi, S., Molyneaux, H. & Ghorbani, A., ''A review of Machine Learning (ML)-based IoT security in healthcare: A dataset perspective,'' *Computer Communications*, vol. 213, pp. 61-77, 2024.

[12]. Millwood, O., Miskelly, J., Yang, B., Gope, P., Kavun, E. & Lin, C., ''PUF-Phenotype: A Robust and Noise-Resilient Approach to Aid Group- Based AuthenticationWith DRAM-PUFs Using Machine Learning,'' *IEEE Transactions On Information Forensics And Security*, vol. 18, pp. 2451- 2465, 2023.

[13]. Kathamuthu, N., Chinnamuthu, A., Iruthayanathan, N., Ramachandran, M. & Gandomi, A., ''Deep Q-learning-based neural network with privacy preservation method for secure data transmission in internet of things (IoT) healthcare application,'' *Electronics*, vol. 11, no. 1, pp. 157, 2022.

[14]. Kumar, R., Joshi, G., Chauhan, A., Singh, A. & Rao, A., ''A Deep Learning and Channel Sounding Based Data Authentication and QoS Enhancement Mechanism for Massive IoT Networks,'' *Wireless Personal Communications*. vol. 130, no. 4, pp. 2495-2514, 2023.

transparent decision-making processes; interfacing with an interoperable framework for operating on cross-platforms; robustness of Federated Learning (FL) methods designed for privacy-preserving distributed training, and standardized and realistic evaluation environments and datasets. These are important first steps in taking AI for congestion control from models to real-world deployments. Results The hybrid, predictive–adaptive AI-based framework showed enhanced network performance in simulations, including reduced latency, jitter, packet loss and capacity usage. Finally, in addition to demonstrating the disruptive impact of AI for improved IoT network resilience, this work offers a clear trajectory for future advancements. With the proliferation of IoT in various critical areas including healthcare, smart cities and industrial automation, the need for deploying intelligible and reliable congestion control mechanisms for secure, scalable, and high-performing communication infrastructures will become increasingly important.

References

Saranya, T., Sridevi, S., Deisy, C., Chung, T. & Khan, M., ''Performance analysis of machine learning algorithms in intrusion detection system: A review,'' *Procedia Computer Science*, vol. 171, pp. 1251-1260, 2020.

[1]. Zhu, R., Ji, X., Yu, D., Tan, Z., Zhao, L., Li, J. & Xia, X., ''KNN-Based Approximate Outlier Detection Algorithm Over IoT Streaming Data,'' *IEEE Access*, vol. 8, pp. 42749-42759, 2020.

[2]. Wu, D., Jiang, Z., Xie, X., Wei, X., Yu, W. & Li, R., ''LSTM Learning With Bayesian and Gaussian Processing for Anomaly Detection in Industrial IoT,'' *IEEE Transactions On Industrial Informatics*, vol. 16, no. 8, pp. 5244-5253, 2020.

[3]. Abououf, M., Mizouni, R., Singh, S., Otrok, H. & Damiani, E., ''Self- Supervised Online and Lightweight Anomaly and Event Detection for IoT Devices,'' *IEEE Internet Of Things Journal*, vol. 9, no. 24, pp. 25285- 25299, 2022.

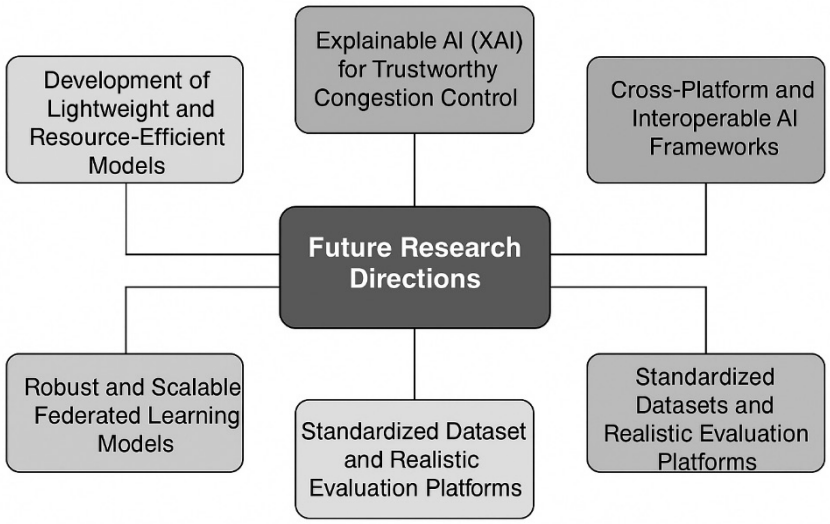


Figure 4: Future Research Directions for AI-Based Congestion Control in IoT Systems

6. Conclusion

The rising size, complexity and heterogeneity of contemporary Internet of Things (IoT) environments have made the traditional congestion control methods inadequate to sustain the network efficiency and responsiveness. In this paper, we conduct a thorough examination of AI-based techniques (i.e., ML, DL, RL, and hybrid techniques) as potential solutions for proactive congestion alleviation. We provide detailed comparisons and show that, although each AI method has its own superiority (e.g., high adaptability in RL, temporal modeling in DL), challenges such as the overwhelming computational cost, lack of interpretability, and poor cross-scenario generalization still exist in terms of STS-IPM. To this end, we outline a number of critical research directions: lightweight and energy-efficient models adapted to edge deployment; integration of Explainable AI (XAI) for

To address this, the research community needs to invest in large, open, diverse, and well-labeled IoT congestion datasets that span different domains (e.g., smart cities, industrial IoT, or healthcare) [45]. Moreover, model behavior should also be validated in realistic testbeds such as digital twins, emulated environments or real-time IoT testbeds, under dynamic and uncertain network scenarios. Projects like IoTBench, FIT IoT-LAB, and EdgeNet are providing encouraging starting platforms for experimentation standardization [46]. All these avenues provide the potential to construct more efficient, transparent and trustworthy AI-based systems that can be functional across the vast and dynamic IoT world. Chasing these constructs will ease the move towards operational research on ICT beyond mere theoretical building blocks towards systems that can actually cope with congestion in future intelligent environments.

We summarize the future research directions covered in this section with a visual summary. We summarize five fundamental directions for promoting AI-based congestion control in IoT systems in Figure 4. These directions cover s both technical optimizations — e.g. Weight and Federated Model developments — and more wider considerations, such as Interpretability, Interoperability and Benchmarking. Between the two, they present a blueprint on how to construct scalable, secure and generalisable AI solutions that can be put in place successfully in diverse IoT settings.

5.4 Robust and Scalable Federated Learning Models

Federated Learning (FL) has recently emerged as a potential privacy-preserving mechanism through which AI models can be trained without the need for centralized data aggregation. In the case of congestion control for IoT, FL could become an interesting alternative, as it enables edge devices to work together to train models whilst keeping data locally. Nevertheless, the issues of the heterogeneity of the data distribution, the model synchronization, and the communication overhead still exist.

In the future work, it is necessary to construct the lightweight FL architectures that are more scalable in the thousands of distributed IoT devices, with a communication-tolerant capability, and with the self-adaptation towards the device-dependent resource constraints [43]. Improvements such as federated averaging with differential privacy, compression algorithm and learning rate adaptation can aid the performance of the proposed framework under non-IID settings which are common in the IoT networks in reality [44]

5.5 Standardized Datasets and Realistic Evaluation Platforms

The lack of standard, publicly available datasets for benchmarking is a fundamental roadblock to advancement of research in AI congestion control. Most previous studies are based on small-scale or synthetic datasets which are not representative for real-world one in terms of traffic dynamics, device mobility, and multi-application interference. Consequently, it is challenging to compare methods across studies, and reproducibility may be lacking.

tend to lack interpretability. This ''black-box" property restricts their acceptance in safety-critical IoT industries including healthcare, autonomous driving, and smart infrastructure where the reasoning of the decisions is vital.

To mitigate this, we need to bring Explainable AI (XAI) approaches inside congestion control frameworks. Techniques such as SHAP [39], LIME [35], or attention mechanisms can be useful for visualizing which features or behaviors contributed to a particular decision. In addressing interpretability, XAI enhances the trust in a system making decisions and facilitates regulatory compliance, in the cases where standards (e.g. GDPR or ISO/IEC 27001 [40]) can be met.

5.3 Cross-Platform and Interoperable AI Frameworks

Asset 1 An ongoing bottleneck in the AI research space is the absence of platform-independent AI models. Most existing approaches are closely bound to a given dataset, hardware platform, or communication protocol. This limits their mobility and prevent them to be deployed in diversified IoT environments, where devices and systems' interoperability are essential.

Future works can also expend in architecting cross-platform AI frameworks to minimize hardware dependency and support multiple IoT standards like CoAP, MQTT, and LoRaWAN [41]. Portability on heterogeneous edge devices can be facilitated using single packaged technologies via containerization (e.g., Docker) and cross-compilation toolchains. Furthermore, use of open APIs and compliance with global communication and data interchange standards will guarantee long-term sustainability and facilitate integration [42].

5. Future Research Directions

To overcome the limitations and research gaps discussed in the preceding sections, the future of AI-based congestion control in IoT systems must focus on building intelligent, efficient, and scalable solutions. This section outlines five critical research directions that aim to improve the performance, adaptability, and trustworthiness of AI-driven congestion management frameworks in real-world IoT deployments.

5.1 Development of Lightweight and Resource-Efficient Models

More IoT devices are resource-constrained in their energy, memory and computation capabilities. Most existing AI models—particularly deep learning networks—are resource-hungry and are not deployable to the edge. So, one of the most promising research directions is to make lightweight and energy-efficient AI models which can locally perform inference tasks with good effect of accuracy.

Recent advances in TinyML and model quantization offer promising opportunities to downsize the model and save memory consumption while maintaining the key performance metrics [37]. Furthermore, pruning, knowledge distillation, and neural architecture search (NAS) can be leveraged to design models tailored to embedded hardware. Further investigation of these approaches can result in low-power, and bandwidth-constrained AI systems, suitable for the IoT platform [38].

5.2 Explainable AI (XAI) for Trustworthy Congestion Control

Though AI, especially deep and reinforcement learning models, has been proved to effectively handle congestion detection and control, they

dedicated to traffic prediction or energy consumption, however few studies provide an end-to-end solution joining all the topics in a coherent way. Figures 3: The Research Stack of the IoT Fragmentation between IoT Layers and Research Requirements Figure 3 shows how common Guides IoT Stack in between the Research Trends and the Layers of the IoT Stack research directions

are map to the corresponding layers in the IoT stack, how it is split into its own research stack pieces, and highlights the need for full-stack, interdisciplinary frameworks FIGURE 3.

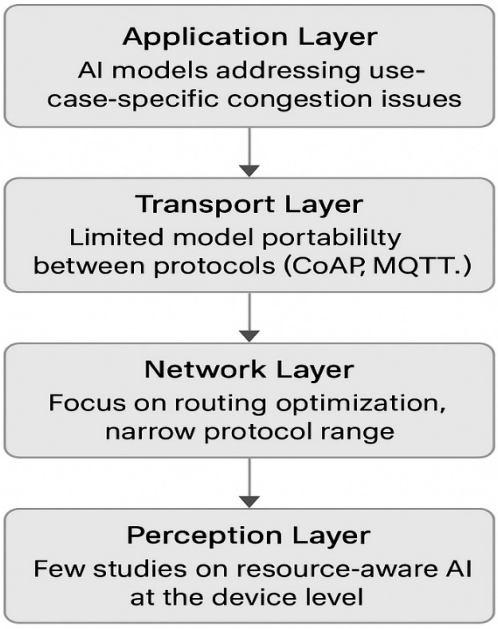


Figure 4.2: Fragmentation of AI-Based Congestion Control Research Across tthe IoT Stack

Figure 3: Fragmentation of AI-Based Congestion Control Research Across the IoT Stack

Table 4: Commonly Used vs. Missing Evaluation Metrics in Congestion Control Studies

Metric Category	Common Metrics	Often Missing or Overlooked Metrics
Performance	Latency, Throughput, Packet Loss	Scalability, Stability under bursty load
Resource Efficiency	—	Energy Consumption, Model Complexity
Security & Trust	—	Resilience to Adversarial Attacks, Explainability
Deployment Readiness	Simulation Results	Real-world Testing, Dataset Generalizability

4.4 Fragmentation in Existing Research Approaches

Existing research work on AI-enabled congestion control is very fragmented, which mostly focus on certain aspects such as routing optimization, anomaly detection, and buffer management separately. However, this step-by-step approach does not take into account the overall nature of congestion in IoT networks –where sensing and transmission layers cooperates each other in more complex manner [33].

Also, cross-fertilization across different domain (e.g., cyber security, embedded systems, network protocol design) is limited. For instance, very little similar research work use explainable AI (XAI) methods to render model decisions interpretable and trustworthy in safety-critical applications [34]. Said lack of cooperation between academia and industry further results in a disconnect between theoretical progress and practical deployable solutions. Information and communications integration is still an unresolved challenge in the development of scalable and resilient AI-based frameworks for IoT congestion control [35], [36]. AI-based congestion control research works are generally not well-organized within different layers of the IoT architecture. Some concentrate on packet routing, while others are

4.3 Limitations in Evaluation Metrics and Experimentation

One of the bottlenecks in the development of AI Fines lies in the absence of commonly accepted benchmarks and evaluation methodologies to enable the fair comparison of AI-based congestion control solutions. Common performance metrics like packet loss, average delay, and throughput provide partial information, but do not fully reflect on important performance aspects in edge learning, such as energy efficiency, fairness in distributed resource allocation, computational load, and model scalability [30], [31].

Besides, most of the previous works have been validated using only the chain of simulation tools (NS-2, NS-3, or OMNeT++). Although experiments can be used to control parameters and the platform design from the ground-up, they may not be able to accurately mimic the simulation. As a consequence, the reliability and reproducibility of such models applied in living tissues is unclear. Furthermore, there is an evident lack of open-source IoT datasets dedicated to congestion analysis, hindering the benchmarking and cross-validation across different methods [32]. Although there have been many publications on performance enhancements created by AI-based congestion control models, a deeper look has uncovered discrepancies (and absences) in the performance evaluation criteria. Most works concentrate on simple metrics (e.g., delay, packet loss) without considering some significant factors such as the efficiency of energy, robust security, and the feasibility of deployment. Table II compares some well-reported measures with those that are frequently unreported or underexplored in the literature.

work effectively in V2X context because of varying mobility, latency, and communication nature [27]. Furthermore, dynamic factors (like node mobility, intermittent connectivity, or protocol heterogeneity) are seldom considered by the current models, even if occurring in realistic IoT scenarios [28]. Moreover, this rigidity drastically restricts model reusability and deployment on a large scale on mixed-use IoT ecosystems [29].

While some congestion control AI models demonstrate promising performance in controlled environments, the performance of others is not easily generalizable to a diverse range of IoT deployments. This problem is due to the heterogeneity in network topologies, node mobility, communication protocols, and application needs. To further characterize this performance gap, Figure 2 presents model performance on representative IoT domains including but not limited to smart home, industrial network, and vehicular network, where it is clear that domain-specific models usually fail when transferred to unseen IoT environments.

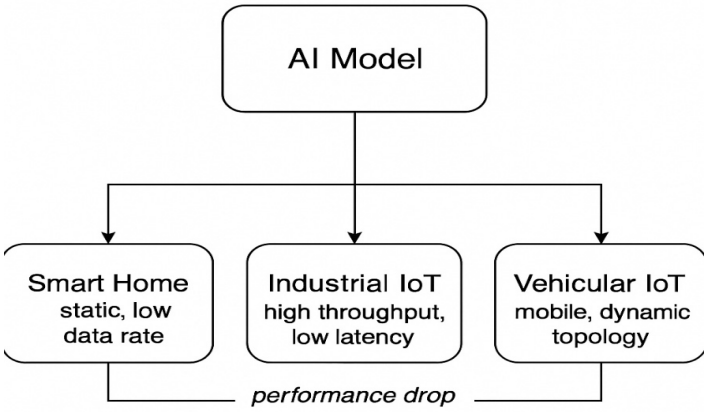


Figure 4.1 Generalization Challenges of AI Models Across Diverse IoT Environments

Figure 2: Generalization Challenges of AI Models Across Diverse IoT Environments

To have a larger scope of the trade-offs among various AI approaches over congestion control in IoT networks, we firstly give a comparison summary. Both approaches have their own advantages depending on the network settings and use case, however, they are not free from scalability, adaptability or resource cost related issues. This table specifies the main strengths and weaknesses of prevalent AI techniques and some selected references extracted from recent work can be found in the following table.

Table 3: Comparative Analysis of AI Techniques for Congestion Control in IoT

AI Technique	Strengths	Weaknesses	References
Machine Learning (SVM, DT)	Interpretable, fast inference	Limited adaptability, requires labeled data	[13], [14], [15]
Deep Learning (LSTM, CNN)	High accuracy, pattern recognition	High resource usage, black-box models	[16], [17], [18]
Reinforcement Learning	Online learning, adaptive	Slow convergence, tuning complexity	[19], [20]
Fuzzy Logic	Handles imprecision, rule-based reasoning	Needs domain expertise for rule crafting	[21], [22]
Hybrid Models	Combines strengths of multiple techniques	Increased algorithmic complexity	[23], [24]

4.2 Gaps in Adaptability and Generalization Across Scenarios

A major shortcoming of existing AI inspired congestion control models is that they are not well adjustable to different network conditions and topologies. Most of the models are designed based on static simulation frameworks or synthetic datasets, under which the variability in real-life IoT scenario does not exist [25, 26].

These models tend to be scenario-specific and do not generalize across network types, domains of application, or device capabilities. For instance, a congestion control model developed in smart home may not

not only varying strengths and weaknesses but also critical research gaps that limit their practical deployment. This section explores these gaps in detail.

4.1 Summary of Strengths and Weaknesses of AI Techniques

Many AI methods have been employed in congestion control such as ML (Machine Learning), DL (Deep Learning), RL (Reinforcement Learning) and Fuzzy logic etc. Each of these approaches provides a different trade-off between performance, interpretability and resource requirement. For example, classical ML algorithms such as SVM and Decision Tree are favored in terms of their interpretability and user-friendly attribute [13], [14]. But in dynamic traffic scenarios, they usually do not perform well due to their reliance on labeled training data and no adaptability [15]. Deep Learning models such as LSTM and CNN, however, have emerged as strong models that capture temporal and spatial patterns and are thus appropriate for transportation anomaly and congestion event prediction [16], [17]. However, their high computational cost and demand for huge amounts of data make them less appropriate to be deployed by resource limited IoT nodes [18].

Reinforcement Learning (RL) algorithm, including Q-learning or Deep Q-Network (DQN), can provide a strong adaption capability by learning the optimal routing or rate-control policies with environmental interaction [19]. However, they usually converge slowly and perform unstably in complex and large-scale topologies [20]. Fuzzy logic and hyb rid AI techniques have been considered to address uncertainty and combine rule based reasoning with learning [21], [22], however, they add complexity and still require domain kno wledge for rule induction [23]. Such trade-offs also motivate the importance of balanced or hybrid architectures with the capacity to exploit the advantages of various AI paradigms [24].

The effective application of AI-enabling congestion control efforts to IoT networks faces a number of technical and practical challenges. These obstacles range from scalability and resource limitations, to timely reaction, security challenges and the lack of interoperability. These challenges in turn have different effects on system performance, reliability and generalisation. Dealing with them well in practice is challenging and requires a mix of architectural changes, lightweight learning and strong security mechanisms. Table 3.1 gives a concise summary of these challenges, that is their effects on IoT network performance, as well as possible mitigation procedures, based on current research directions and best practice Table 3.

Table 2: Summary of Challenges in AI-Based Congestion Control for IoT

Challenge	Impact	Potential Mitigation
Scalability in large networks	High communication cost, model overhead	Decentralized learning, hierarchical models
Device constraints	Energy drain, infeasible model deployment	Lightweight models, TinyML, quantization
Real-time adaptability	Delayed decisions, poor responsiveness	Online learning, edge inference
Security & privacy risks	Data manipulation, adversarial behavior	Secure model training, anomaly detection, encryption
Generalizability across IoT systems	Limited reusability and deployment flexibility	Modular, protocol-agnostic architectures

4. Comparative Analysis and Research Gaps

Artificial Intelligence (AI) has emerged as a powerful solution for managing network congestion in IoT environments, offering predictive and adaptive capabilities far superior to traditional rule-based methods. However, an in-depth comparison of existing AI-driven congestion control techniques reveals

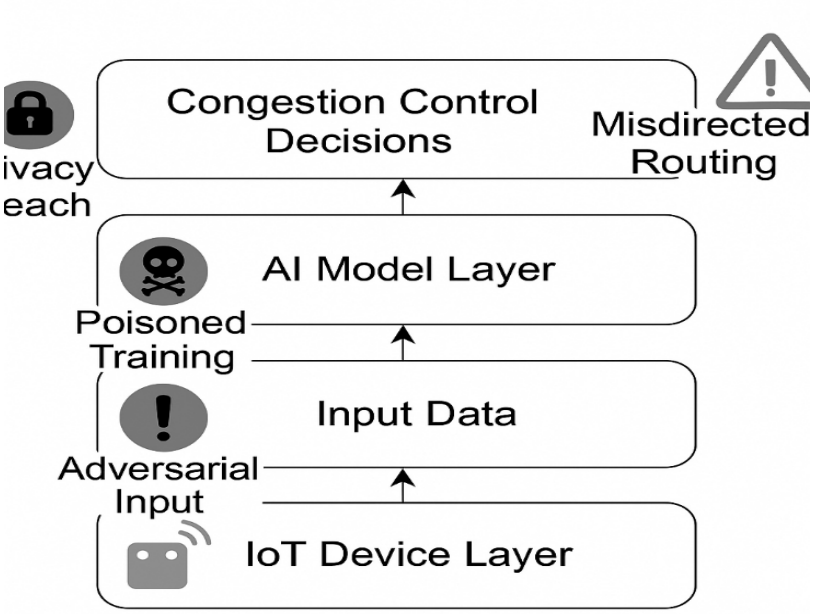


Figure 1. Threat Landscape in AI-Based IoT Congestion Control Systems

3.5 Lack of Generalizability and Cross-Platform Support

The majority of congestion avoidance AI approaches in the IoT context lack of a strong generalization and interoperability capabilities. These models are usually developed and trained for given datasets, specific network topologies, and a limited class of communication protocols. Therefore, their performance becomes remarkably decreased when utilized in a variety of IoT networks (e.g., including different supporting type of devices, network architectures, and application domains). Furthermore, most solutions lack portability for different IoT operating systems and hardware platforms, which makes cross-deployment difficult. This fragmentation is a major obstacle for mass deployment, because typical IoT scenarios require the standardized, adaptable and platform-agnostic systems which work across heterogeneous environments [11].

centralized collection and processing of IoT data (made possible by the data abundancy and the high–capacity storage and analytic systems) leads to potential privacy leakage [2]: users' privacy is threatened when sensitive personal data -such as their location or environmental and traffic flows – can be inferred or intercepted. E.g., an attacker can modify the packet transmission rates to make the network send false congestion notifications, leading data to be redirected over unsafe or compromised paths. To prevent such risks, we need to design secure and privacy-preserving AI systems that embrace secure learning, anomaly detection, differential privacy, and federated learning to ensure the integrity of AI models and the privacy of users [12].

To gain deeper understanding about the various threat vectors in the AI-implementation into IoT congestion control, a layered threat model is introduced. A similar model is presented in [20] that describes different levels of AI-informed decision-making, but does not emphasize how each layer of the decision-making pipeline is subject to its own specific family of attack. At the input layer, one can generate adversarial examples to cause the AI model to make wrong predictions. In the learning data, poisoning attacks can disrupt the learning process using manipulated data in the training phase. Even more privacy leakage will happen in centralized data processing when personal behavior and content is uncovered. Last, faulty or compromised AI outputs can result in misguided routing decisions, possibly posing a threat to service reliability and security. These vulnerabilities underscore the importance of secure, interpretable, and robust AI algorithms for IoT network management. Fig. 1 The threat landscape of AI-based IoT congestion control systems

require significant memory, CPU/GPU power, and energy, which are often unavailable on edge devices. Common limitations include limited battery life, a small memory footprint, and no hardware acceleration, as basic IoT chips lack support for GPUs or AI co-processors. For example, a convolutional neural network model for congestion classification may require 20-30 MB of RAM and a GPU, making it incompatible for agricultural applications.

3.3 Real-Time Adaptability and Model Latency

AI models, particularly those using supervised learning, face challenges in real-time response to network traffic fluctuations due to high inference latency, model drift, and delayed decision-making. These issues can be addressed through reinforcement learning and edge AI, which allow local adaptation and distributed decision-making, thereby reducing the need for constant offline training and periodic updates.

3.4 Security and Privacy Concerns in AI-Powered Systems

The union of AI and IoT for congestion control presents a set of novel security and privacy threats that go beyond those of traditional networking. Specifically, AI models, especially congestion prediction/control ones, could potentially be attacked via adversarial traffic patterns, utilizing subtle noise in the input traffic features to fool the classifier or regressor to make wrong or misleading decisions. In addition, on the training phase, data poisoning attacks represent a serious threat, as an adversary may inject misguided or malicious data in the model in order to change the AI behavior, thus causing the AI model to operate in a nonoptimal way, or to generate fake congestion events. Moreover, recent results have highlighted the fact that the

effectiveness, scalability, and sustainability of AI-based congestion control frameworks.

3.1 Scalability Challenges in Large IoT Networks

The networks here will be much denser, reaching up to millions of heterogenous devices, and scaling AI models here poses direct challenge. Much of proposed AI solutions are tested and evaluated Using low scale Testbeds or simulated environments that do not reflect the real-world challenges and dynamics of such large scale IoT systems. Scalability issues to address include: a) Scale of users: With the increasing popularity of key brewkins over the internet domain, supporting the scale of the brewkins themselves may become an issue the final implementation will need to address.

a) Communication Overhead: As the number of messages between devices and central/cloud-based AI modules increases, the bandwidth will eventually become saturated.
b) Complexity in coordination: Synchronizing model update or decision in distributed agents is made-adifficult.
c) Model Explosion: With additional devices, the number of achievable states and actions also increases, making the models progressively larger and slower. Example: In a smart city scenario, employing a single global model to predict congestion in 500,000 traffic sensors can delay the prediction and create bottlenecks by aggregating the data centrally.

3.2 Computational and Energy Constraints in IoT Devices

IoT devices are typically low-power, lightweight systems designed for specific tasks. AI models, such as Deep Learning and Reinforcement Learning,

Table 1: Comparative Analysis of AI Approaches for Network Congestion Mitigation in IoT Systems

AI Technique	Key Application Area	Advantages	Limitations	References
Machine Learning (SVM, Decision Trees)	Traffic prediction, routing decisions	Simple implementation, interpretable models	Limited adaptability, requires labeled data	[1], [2]
Reinforcement Learning (Q-learning, DQN)	Dynamic routing, real-time congestion control	Self-adaptive, no prior data required	Slow convergence, high computational overhead	[3], [4]
Deep Learning (LSTM, CNN)	Temporal/spatial pattern recognition	High accuracy, capable of complex feature extraction	Needs large datasets and resources	[5], [6]
Fuzzy Logic	Congestion prediction under uncertainty	Handles imprecise data, rule-based reasoning	Requires expert knowledge for rule creation	[7]
Hybrid Models (Fuzzy + GA/ML)	Optimization of transmission parameters	Improved performance, adaptable	Increased algorithmic complexity, harder to tune	[8]
Federated Learning & Edge AI	Distributed learning, privacy-preserving	Scalable, data privacy, low latency at edge	Lack of standards, interoperability challenges	[9], [10]

3. Challenges and Issues in AI-Based Congestion Control

Artificial Intelligence (AI) techniques provide dynamic and intelligent capabilities for mitigating network congestion in IoT environments. However, real-world deployment of AI-based solutions still faces several technical and operational barriers. This section analyzes the core challenges that limit the

Edge AI has also been developed as an alternative approach by moving the AI computation near to the IoT devices to decrease the latency and save the bandwidth. Systems can process data locally using lightweight AI models without having to round-trip to the cloud riding in the fast lane, reducing response times. Edge-based AI systems like [9] have demonstrated increased network efficiency and reduced dependence on cloud-centric resources. However, issues in model deployment, security, and model updates are still some problems need more investigation in edge AI research [10].

However, there still exist a number of research gaps. However, many AI driven mechanisms do not provide general solutions across IoT architectures and operate on specific cases or protocols. Further, the incorporation of explainability to AI models for network management is still low, leading to trust and adoption. Furthermore, there are requirements for light weight and energy efficient AI models that are deployable in real time on resource constrained devices. For next-step research, we recommend that researcher should pay great attention to design a scalable, interpretable, and cross-compatible AI framework that can self-adapt into all kind of fluctuating IoT network scenarios.

adapting transmission rates and routing decisions to find the shortest path and reduce packet loss. Q-learning and Deep Q-Networks Q-Networks (DQN) have been proven to be work well in reacting to dynamic traffic by not defining any fixed rules [3]. For instance, [4] used RL to load balance the IoT nodes leading to decreases in both end-to-end delay and energy consumption. Nevertheless, RL still faces challenges for convergence speed and scalability in large-scale networks.

Fuzzy Logic and Hybrid AI models are another category of solutions designed to handle uncertainty and imprecision characteristic of IoT data delivery. Ambiguous input and human-like response can be modeled by fuzzy system used for prediction of congestion. In conjunction with ML or evolutionary algorithms, these hybrid approaches have proven to be that promising in the context of heterogeneous traffic [5]. For example, under dense IoT deployments, [6] proposed a Fuzzy-Genetic Algorithm for adaptive traffic rerouting. However, these approaches can be computationally complex for real-time decision making.

Another popular direction is the use of Deep Learning paradigms like Long Short Term Memory (LSTM) and Convolutional Neural Networks (CNNs) for the temporal and spatial congestion detection. The LSTM and CNN offer good time-series and spatial learning for predicting traffic bursts in LTE core networks and spatial relationships in mesh IoT topologies, respectively. These models have been used with the application of anomaly detection, traffic classification and congestion prediction with high precision [7]. Nevertheless, DL models usually need abundant training samples and computing resources and may not be applicable to resource-constrained IoT devices [8].

b) Section 3 presents the research methodology, dataset, feature extraction method, and model building.
c) Experimental setup and performance evaluation results are described in section 4.
d) Section 5 presents the developed hybrid AI framework, describing its architecture and realization.
e) Section 6 summarizes findings, discusses limitations, and sets future work and concludes the paper.

By providing a holistic and intelligent congestion control guidance, this work helps push forward resilient, adaptive, and scalable IoT networking systems, tackling one of the core obstacles of an universal deployment of IoT infrastructures.

2. Literature Review

Explosive growth of Internet of Things (IoT) devices led to numerous challenges in network management – congestion control is one of the major concerns. Traditional congestion window updating and the packet dropping policies are not adapted to dynamic and heterogeneous IoT environments. As a result, AI has come into consideration as a prospective solution that can provide decision-making in real-time, and predict future events, which are both requirements for a proper IoT system. In particular, Machine Learning (ML) and Deep Learning (DL) models have been exploited in predicting the traffic pattern, optimizing routing paths and acting proactively in handling data loading to improve QoS [1], [2].

RL has attracted much attention in real-time congestion control in IoT. RL models can optimize policies based on environmental interactions,

routing optimization or anomaly detection, and rarely build a complete framework that proactively and comprehensively tackles the congestion issues.

1.1. Objectives of the Study

This paper attempts to bridge this gap by proposing an AI integrated hybrid congestion control model for IoT networks. The specific aims of the study are to:

a) Understand which types of congestion in different IoT deployment environments, and what causes of such congestion.

b) Examine and contrast the current AI-based strategies employed in congestion prediction and control, respective of performance/ success under various network scenarios.

c) Develop a hybrid AI model to integrate predictive intelligence (e.g., predictive learning), with adaptive decision-making (e.g., reinforcement learning) for effective and scalable congestion alleviation.

d) Verify the proposed model through simulations and experiments on real-life IoT data sets and the system performance evaluated in terms of latency, packet loss rate, throughput, energy consumption with all relevant data.

1.2. Paper Organization

The rest of the paper is structured as follows:

a) In Section 2, the related works have been studied for both conventional congestion control techniques, and AI techniques from which congestion control has been obtained in IoTcitation.

static routing procedures and Active Queue Management (AQM). Although these approaches are effective in traditional networking scenarios, they are in most cases not compatible with IoT scenarios as they react to situations rather than prevent potential impending events, lack of multi-level adaptation mechanisms, and cannot deal with the heterogeneity and dynamic IoT traffic patterns [2]. These legacy methods monitoring and reacting to congestion after it has been made when it is too late, resulting in reduced system performance, especially when there are very dynamic topologies or IoT domains with mobile moving elements.

To address these challenges, Artificial Intelligence (AI) has become the hottest paradigm for intelligent network management. To achieve such functionality, AI methodology, particularly those deriving from Machine Learning (ML), Deep Learning (DL) and Reinforcement Learning (RL) have provided real time traffic analysis, congestion prediction and autonomous decision making [3]. In contrast to static algorithms that simply react to congestion after it occurs, AI can learn from traffic data over time, identify changing patterns, and flexibly and dynamically assign network resources to preemptively address or contain congestion before it has a significant impact on service quality. For example, reinforcement learning agents can optimise routing decision or bandwidth allocation according to responses from environment, and deep learning models can forecast the congestion zone using spatio-temporal features of traffic.

Although AI has great potential to address network congestion, issues remain in terms of appropriate model selection, computational efficiency, and trade-off between prediction accuracy and resource limitations. What's more, previous research works tend to consider a set of partial viewpoints, e.g.,

1. Introduction

The Internet of Things (IoT) is poised to revolutionize the communication infrastructure as we know it by interconnecting billions of smart devices – from wearable health monitors and smart home appliances to industrial sensors and autonomous vehicles – in a gigantic communication network which can sense, process, and exchange information in real time. This hyperconnected environment allows for advanced applications in areas ranging from healthcare and transportation to smart cities and industrial automation, fostering data-driven decision-making, predictive analytics and autonomic contro51. In industry forecasts, it is estimated the number of IoT-connected devices could reach over 30 billion by 2030, with large scale data produced, leading to unprecedented requirements for communication networks and computational infrastructures [1].

Although the expansion of IoT offers values in innovation and convenience, it also poses a series of new networking problems: network congestion is the most severe one. The congestion of a network is created when not all the data sent are handled, and this results in delay, packets loss, buffer overflow, and reduces in QoS. When delay becomes a factor for applications that are sensitive to processing latency - for instance, remote surgery, or self-driving cars - the communication delay can be the key difference between success and disaster. Besides, as IoT devices is tend to operate in resource-scarce environments, e.g., under low power, processing and bandwidth, managing network traffic effectively becomes an essential consideration to guarantee the scalability, reliability and responsiveness of IoT systems.

Generally, rule-based mechanisms have been used in both control algorithms, eg., Transmission Control Protocol (TCP) variants, only some

Abstract

The unprecedented explosion of Internet of Things (IOT) devices has elevated the requirements of the network infrastructures to unprecedented levels, causing severe congestion problems, especially in applications which demand low latency, high throughput, and real-time feedback. Static routing protocols, AQM, and TCP variants are some of the traditional mechanisms for congestion control that are unable to perform efficiently in dynamic and diverse IoT environments as they are reactive-based and inflexible. To this end, in this paper, we explore the promising ability of Artificial Intelligence (AI) methods such as Machine Learning (ML), Deep Learning (DL), Reinforcement Learning (RL), and their combination in natura for proactive and intelligent traffic management for IoT. A comparative review of strengths (e.g., adaptivity in RL, pattern recognition in DL) and weaknesses (in terms of its scalability, interpretability, resources) of each method is also discussed. Moreover, the paper indicates some crucial research challenges on model generalization, evaluation criterion and platform integration. Future possible research directions to bridge these gaps include the development of lightweight AI architectures, Explainable AI (XAI) frameworks, cross-platform model deployment, scalable FL, and standardized benchmarking datasets. This work also leads to a hybrid AI model for traffic congestion prediction and control with an application of simulation tool and real data. Simulation results show significant improvements in latency, packet loss, and energy consumption. Finally, the study presents a ground work for incorporating the scalable, secure and intelligent AI enabled congestion control systems in a wide area of IoT applications.

Keywords: - Internet of Things (IoT), Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Reinforcement Learning (RL), Federated Learning (FL), Explainable Artificial Intelligence (XAI)

Artificial Intelligence Approaches to Mitigating Network Congestion in IoT Systems / Original article

Aysar Hadi Oleiwi
Department of Computer and Electrical Engineering ,
College of Engineering ,University of Tabriz, Iran
ayserhadi7@gmail.com
Orcid=0009-0003-2264-9919

[14]. N. Sharma and Dr. A. V Krishna Prasad, (2021),"File-level deduplication by using text files – Hive integration", in Proceedings of the International Conference on Computer Communication and Informatics (ICCCI), p. 6.

[15]. G. Sujatha and Dr. Jeberson Retna Raj, (2021),"Improving the efficiency of deduplication process by dedicated hash table for each digital data type in cloud storage system", Webology, Vol. 18, Special Issue on Artificial Intelligence in Cloud Computing, pp. 288-301.

[16]. N. A. Jaafar, (2024),"A study on improving the accuracy and effectiveness of similarity detection processes in text files using NLP techniques", Volume 6, Issue 9, Al-Esraa University College Journal for Engineering Sciences.

[17]. S. A. Talib, (2024),"The importance of cryptography in cloud computing", Volume 6, Issue 10, Al-Esraa University College Journal for Engineering Sciences.

[18]. Ammar Zakzouk, Bassim Oumran, and Hasan Hasan, (2024),"ALLI: a high-performance approach to data deduplication in Hadoop using enhanced hashing and two-level indexing techniques", International Journal of Performability Engineering, Vol. 20, No. 12.

8. References

[1]. Y. Himeur, H. Elouadrhiri, L. Khoukhi, and N. Beni-Hssane, (2023),"AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives", Artificial Intelligence Review, Vol. 56, No. 6, pp. 4929-5021.

[2]. M. Akbar, *et al.*, (2024),"Enhanced authentication for de-duplication of big data on cloud storage system using machine learning approach", Cluster Computing, Vol. 27, No. 3, pp. 3683-3702.

[3]. M. Muniswamaiah, T. Agerwala, and C. Tappert, (2019),"Big data in cloud computing review and opportunities", International Journal of Computer Science & Information Technology (IJCSIT), Vol. 11, No. 4, pp. 43-57.

[4]. V. Schmitt and J. Jordaan, (2013),"Establishing the validity of MD5 and SHA-1 hashing in digital forensic practice in light of recent research demonstrating cryptographic weaknesses in these algorithms", International Journal of Computer Applications, Vol. 68, No. 23, pp. 40-43.

[5]. M. Eichlseder, F. Mendel, and M. Schläffer, (2014),"Branching heuristics in differential collision search with applications to SHA-512", in Proceedings of the Fast Software Encryption - FSE 2014, p. 16.

[6]. M. Kathiravan, *et al.*, (2023),"A cloud based improved file handling and duplicate removal using MD5", in 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), IEEE.

[7]. Wikipedia, (2023),"Audio file format", available at: https://en.wikipedia.org/wiki/Audio_file_format

[8]. N. A. Naveen and V. Ravi, (2015),"Client-side deduplication scheme for secured data storage in cloud environments", International Journal of Engineering Research & Technology (IJERT), Vol. 4, No. 5, pp. 1465-1467.

[9]. V. S. R. and D. K. Singh, (2016),"Secure deduplication techniques: a study", International Journal of Computer Applications, Vol. 137, No. 8, pp. 41-43, doi: 10.5120/ijca2016908874.

[10]. S. Parth, G. Amit, P. Sandipkumar, and P. Priteshkumar, (2016),"Efficient cross-user client-side data deduplication in Hadoop", Journal of Computers, DOI: 10.17706/jcp.12.4, pp. 362-370.

[11]. I. Vaidya and R. Nath, (2017),"An improved de-duplication technique for small files in Hadoop", International Research Journal of Engineering and Technology (IRJET), Vol. 4, No. 7, pp. 2040-2045.

[12]. R. Hudagi and A. Urabinahatti, (2018),"Efficient deduplication using Hadoop", International Journal of Latest Trends in Engineering and Technology, Vol. 10, No. 3, pp. 236-238.

[13]. W. Zhang, B. Shao, G. Bian, and Q. He, (2020),"Research on multifeature data routing strategy in deduplication", Scientific Programming, Vol. 2020, Article ID 8869237, pp. 11.

formats, leading to improved efficiency and quicker processing times when managing diverse audio formats in a cloud environment.

The proposed technique enhances the search process by reducing the number of comparison operations, leading to faster performance and a lower likelihood of data collisions. Unlike methods that rely on multi-table comparisons—where all files are compared with each other, increasing the chances of similar keys and subsequent collisions—our approach confines comparisons to files within the same format index. This targeted comparison strategy effectively minimizes collision probability while maintaining efficient search speeds.

7. Conclusion

With the increasing volume of data received by the cloud system, which includes duplicate data, it is not enough to design techniques that only focus on efficient duplicate detection. In the face of this massive amount of data, it is necessary to minimize the time required to process duplicate data as much as possible. In this research, we have built a multi-indexing system based on metadata (file format) to reduce the number of comparisons and thus accelerate the search process for the partition key. Reducing the number of comparisons also reduces the probability of collisions between partition keys. In this research, we also used the MD6 algorithm to calculate the hash value for the files, which produces a 512-bit key. The longer the key length, the lower the collision rate. In the future, it is possible to build a duplicate data elimination system that relies on multiple tables and multiple indexes to minimize the time complexity as much as possible.

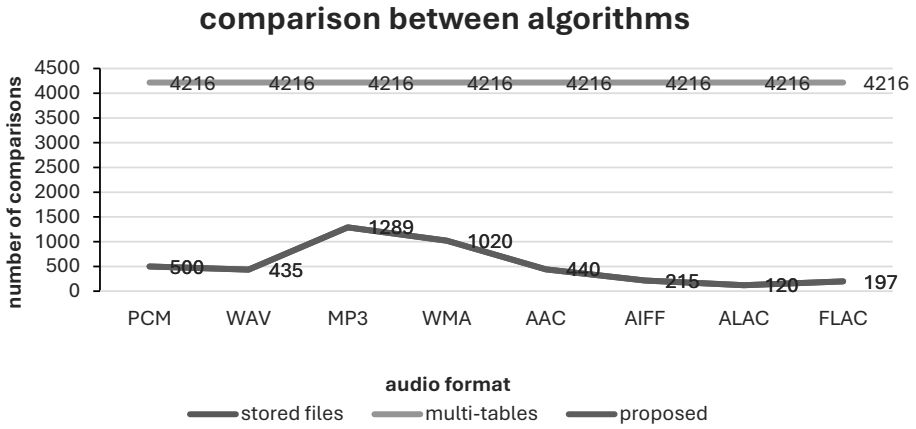


Figure (7): Change in the number of comparisons.

6. Results and Discussion

The multi-table-based technique stores all hash keys of audio files in a centralized manner. To find a match for a specific file, its key must be compared against the keys of all other files in the system. This approach results in a high overall time complexity, as the number of comparisons increases significantly with the total number of audio files. Consequently, the process becomes less efficient, particularly when handling large datasets.

To address this issue, the proposed algorithm intelligently divides the main table of audio files into nine separate indexes, each corresponding to a specific file format. Each index contains only the keys of files that share that particular format. When a new audio file is uploaded, it is assigned a unique key and stored in the index that matches its format. Consequently, when searching for a specific file, the comparison is limited to the relevant format index, drastically reducing the number of comparisons required. This proposed approach eliminates unnecessary comparisons across different

6- 215 audio files in AIFF.

7- 120 audio files in ALAC.

8- 197 audio files in FLAC.

The figure (6) illustrates the difference in the number of comparison operations between the proposed algorithm and the algorithm based on multiple tables, in the case of an audio file of a specific format being received in the cloud.

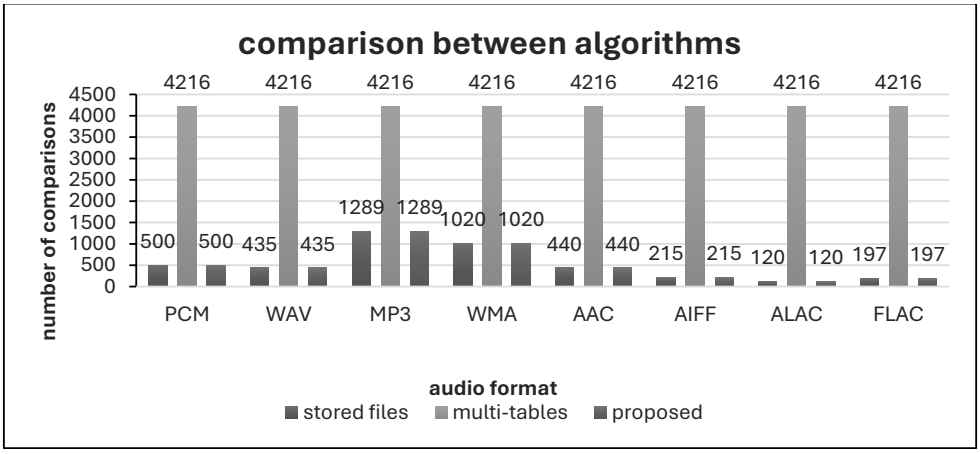


Figure (6): Role of proposed algorithm.

Using the proposed algorithm, the number of comparison operations does not change based on the total number of audio files stored in the cloud. Instead, it depends on the number of files stored for each individual format. For example, let's consider a scenario where there are 500 MP3 files stored in the cloud and no WAV files. If a WAV file is received, it will be automatically stored in the cloud without any comparison operations being performed. This is because there are no other WAV files to compare it with.

The figure (7) illustrates the change in the number of comparisons using the proposed technique.

Based on the table (1), the figure (5) illustrates the effectiveness of the proposed algorithm and compares the results obtained by this algorithm with the algorithm based on multi-table.

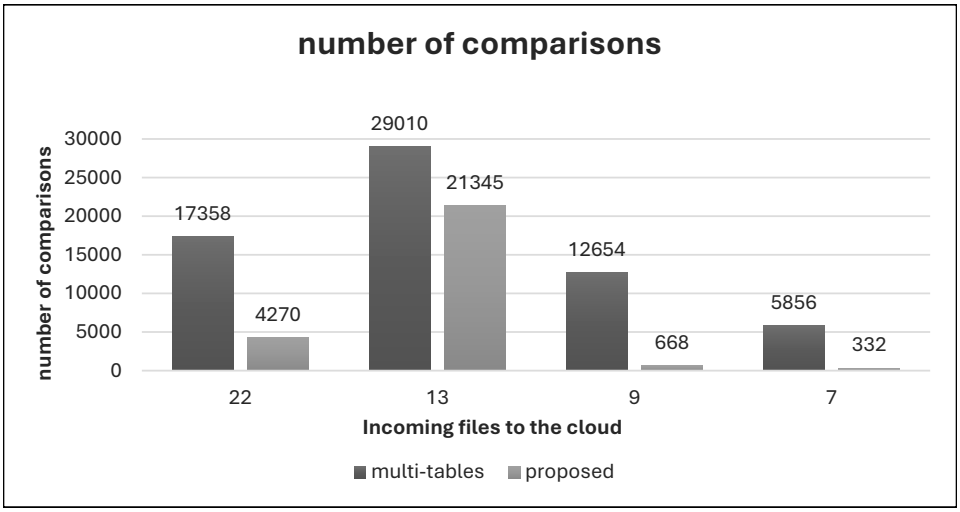


Figure (5): Effectiveness of the proposed algorithm.

From the figure above, it can be observed that the proposed algorithm is more efficient compared to the technique based on multiple tables. With the proposed algorithm, each audio file coming to the cloud will only be compared with other audio files of the same format.

For further clarification, let's consider that the total number of audio files stored in the cloud is 4216, distributed as follows:

1- 500 audio files in PCM.

2- 435 audio files in WAV.

3- 1289 audio files in MP3.

4- 1020 audio files in WMA.

5- 440 audio files in AAC.

Using the proposed algorithm: The number of comparisons for WMA audio files is 450*4= 1800, for WAV audio files is 296*4=1184, and for FLAC audio files is 200*3=600, the total number of comparison operations: 1800+1184+600=3584 comparisons.

Using the proposed based on multi-tables [15]: The hash key of any incoming file to the cloud will be compared with all the hash keys stored in the audio files hash table, and therefore, the number of comparison operations is 1923*11= 21153.

5.4 Other Models

The table (1) illustrates the experimental models that differ in the number of files stored in the cloud, the number of files incoming to the cloud, and the maximum number of required comparison operations. This is achieved using the proposed algorithm and the technique based on multiple tables.

Table (1): Effectiveness of the proposed algorithm.

Experiment	Incoming files to the cloud Multi-tables proposed				Number of audio files in the cloud					Number of Comparisons	
Experiment1	22	PCM	MP3	WAV	PCM	MP3	WAV	FLAC	WMA	17358	4270
		12	6	4	123	415	76	92	83		
Experiment2	13	MP3	AIFF	AAC	MP3	WAV	WMA	AIFF	AAC	29010	21345
		5	4	4	625	120	870	1700	2855		
Experiment3	9	WMV	WMA	AAC	PCM	WMA	FLAC	AIFF	AAC	12654	668
		4	3	2	413	210	31	521	19		
Experiment4	7	MP3	FLAC	WMA	WAV	WMA	AAC	ALAC	AIFF	5856	332
		3	3	1	47	332	498	12	87		

Using the proposed based on multi-tables [15]: The hash key of any incoming file to the cloud will be compared with all the hash keys stored in the audio files hash table, and therefore, the number of comparison operations is 1923*56= 107688.

5.2 Second Model

The number of incoming audio files transmitted to the cloud is 23 audio files, distributed as follows:

1- 9 audio files in MP3 format.
2- 6 audio files in PCM format.
3- 11 audio files in WMA format.
4- 6 audio files in AAC format.

Using the proposed algorithm: The number of comparisons for MP3 audio files is 800*9= 7200, for WMA audio files is 450*11=4950, and there isn't a comparison process for AAC and PCM format. Therefore, the total number of comparison operations: 7200+4950= 12150 comparisons.

Using the proposed based on multi-tables [15]: The hash key of any incoming file to the cloud will be compared with all the hash keys stored in the audio files hash table, and therefore, the number of comparison operations is 1923*23= 44229.

5.3 Third Model

The number of incoming audio files transmitted to the cloud is 11 audio files, distributed as follows:

1- 4 audio files in WMA format.
2- 4 audio files in WAV format.
3- 3 audio files in FLAC format.

5. Experiments

In this section, we will present multiple models that showcase the outcomes of the proposed technique and we will compare the results of the proposed algorithm with the algorithm based on multiple hash table [15]. These models vary in the number of incoming audio files transmitted to the cloud and the file format. The number of audio files stored in the cloud is 1923, distributed as follows:

1- 800 audio files in MP3 format.
2- 450 audio files in WMA format.
3- 296 audio files in WAV format.
4- 200 audio files in FLAC format.
5- 177 audio files in AIFF format.

5.1 First Model

The number of incoming audio files transmitted to the cloud is 56 audio files, distributed as follows:

1- 22 audio files in MP3 format.
2- 17 audio files in WAV format.
3- 11 audio files in WMA format.
4- 6 audio files in AAC format.

Using the proposed algorithm: The number of comparisons for MP3 audio files is 800*22= 17600, for WMA audio files is 450*17= 7650, for WAV audio files is 296*11=3256, AAC files are unique data, and can be stored in the cloud without the need for comparison processes since the cloud system does not store any AAC audio files. Therefore, the total number of comparison operations: 17600+7650+3256= 28506 comparisons.

5. Determine Storage Decision: If the hash value is found in the index, it indicates that the file is a duplicate and should not be stored. If the hash value is not found, it means the file is unique and can be stored in the cloud storage.
6. File Storing: If the file is determined to be unique, it is stored in the cloud storage system.
7. Index Updating: The hash value of the stored file is added to the corresponding index for future reference.

By following these steps, the system can efficiently determine whether to store the uploaded file or not based on its uniqueness and thus, it avoids storing duplicated audio files. The figure (4) shows the general scheme of the proposed technique.

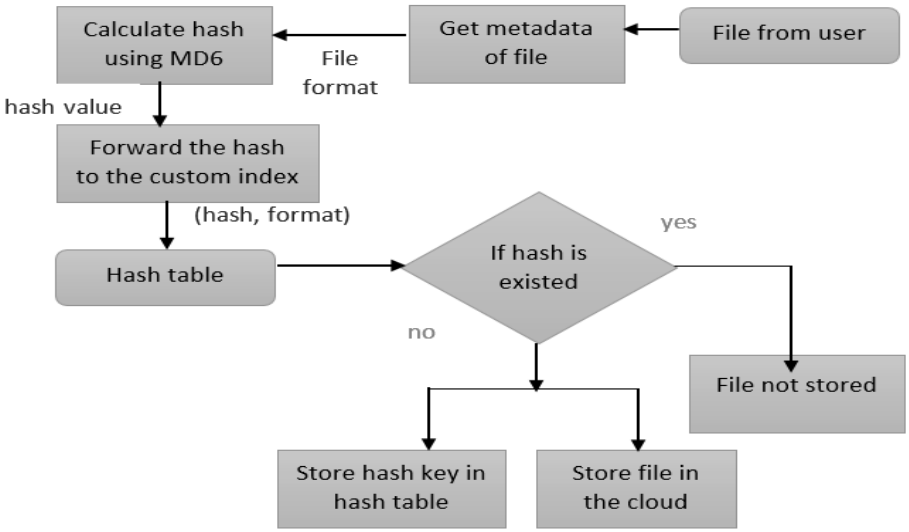
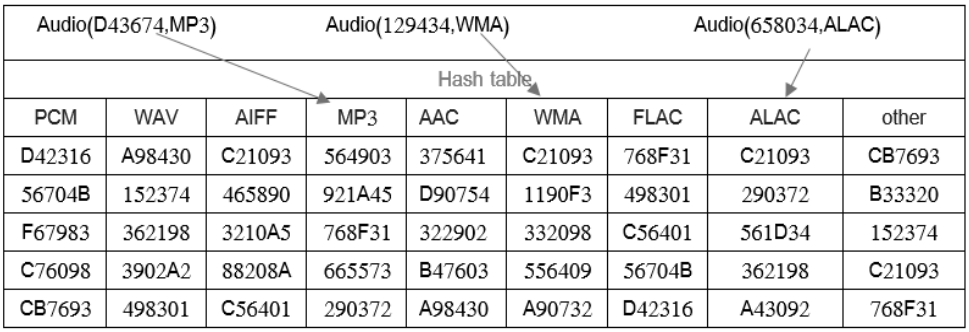


Figure (4): proposed algorithm

Figure (4): Proposed algorithm.

is calculated using the MD6 algorithm. Based on the format of the file, the file's hash value is compared with the hash values stored in the index specific to that format. If the hash value already exists in the corresponding index, it indicates that the file is a duplicate and therefore should not be stored again. However, if the hash value is not found in the index, it means the file is unique and can be stored in the cloud storage. Additionally, the hash value of the file is added to the corresponding index for future reference. The figure (3) shows the routing of hash values.



PCM	WAV	AIFF	MP3	AAC	WMA	FLAC	ALAC	other
D42316	A98430	C21093	564903	375641	C21093	768F31	C21093	CB7693
56704B	152374	465890	921A45	D90754	1190F3	498301	290372	B33320
F67983	362198	3210A5	768F31	322902	332098	C56401	561D34	152374
C76098	3902A2	88208A	665573	B47603	556409	56704B	362198	C21093
CB7693	498301	C56401	290372	A98430	A90732	D42316	A43092	768F31

Figure (3): Routing system.

4.3 Technique Scheme

Based on the provided information, the stages of the technique can be arranged as follows:

1. File Upload: The user uploads the file to the cloud storage system.
2. Read Metadata: The metadata of the uploaded audio file is read to determine its format.
3. Calculate Hash Key: MD6 algorithm is used to calculate the hash key of the file.
4. Index Comparison: Based on the file format, the calculated hash key is compared with the hash values stored in the corresponding index.

4. Proposed Algorithm

We developed an algorithm to deduplicate audio files. The algorithm consists mainly of two phases:

1- building the indexing system.

2- deduplication.

4.1 Index System

Index System consists of a table with multi-indexes. Nine indexes for the most common audio files formats and one index for other formats. The first index is for Pulse-Code Modulation index. The second index is for Advanced Audio Coding. The third index is for Audio Interchange File Format. The fourth index is for MPEG-1 Audio Layer3. The fifth index is for Advanced Audio Coding. The sixth index is for Windows Media Audio and. The seventh index is for Free Lossless Audio Codec. The eighth index is for Apple Lossless Audio Codec. The last index is for other formats. The figure (2) shows the index system and it contains examples for hash values:

Hash table								
Index 1	Index 2	Index 3	Index 4	Index 5	Index 6	Index 7	Index 8	Index 9
PCM	WAV	AIFF	MP3	AAC	WMA	FLAC	ALAC	other
D42316	A98430	C21093	564903	375641	C21093	768F31	C21093	CB7693
56704B	152374	465890	921A45	D90754	1190F3	498301	290372	B33320
F67983	362198	3210A5	768F31	322902	332098	C56401	561D34	152374
C76098	3902A2	88208A	665573	B47603	556409	56704B	362198	C21093

Figure (2): Index system.

4.2 Deduplication

In this approach, when a file is uploaded to the cloud storage, the metadata of the audio file is read to determine its format. Then, a hash key

use in many deferent fields and applications. The results showed that it is possible to accurately determine the similarity between texts [16].

10- In 2024, S. A. Talib, The study begins with the historical background of cryptography and then moves on to analyze trends in information technology before discussing on the challenges to security in cloud computing. The goal of its publication is to highlight the importance of using cryptographic technologies for the protection of information con_dentiality, its integrity and compliance. The combination of literature review and the case study method was applied, pulling knowledge from examples of good cryptographic use and the lessons learned from breaches in security. They reveal how cryptology is crucial in the furthering of cloud security and in maintaining customers' faith. Key information technology practitioners are encouraged not to relax their guard and adopt an integrated security strategy to ward o future attacks. Conclusion of the study is in synthesis of the subsequent analysis, highlighting the continued applicability of cryptography for Cloud Information System security [17].

Thus, the previous studies and their solutions as presented above are not sufficient because one of the parameters in the process of eliminating duplicate data is the time required to implement the technique, which is mainly related to the number of comparison operations for the hash key of the incoming file with the stored keys. The fewer the number of operations, the less the execution time. For any incoming audio file, it will be compared with all stored files (text, images, videos, audio) if a single indexing table is used. However, it will be compared with all audio files if multiple indexing tables are used. This requires a long execution time, which increases with the increase in the stored files.

6- In 2020, Weiqi ZHANG *et al*. proposed a technique for deduplication in Hadoop. Hash table was designed in namenode. For every block of files, a hash key using SHA-512 was used. To determine the uniqueness of each block, the generated hash key is compared to the existing keys in the system. If there is no match, it indicates a unique block, which gets stored in the Hadoop Distributed File System (HDFS). SHA-512 algorithm is used to generate a hash value with 512 bits, which results in a lower collision rate compared to previous algorithms. This ensures a higher level of confidence in identifying and storing unique blocks of data [13].

7- In 2021, Niteesha Sharma and Dr. A. V Krishna Prasad, proposed a technique to solve the storage issues and deduplication in Hadoop. A table with hash keys is built in Hbase and SHA-256 is used. The process of reading from HBase is faster than Hadoop Distributed File System (HDFS) [14].

8- In 2021, G. Sujatha and Dr. Jeberson Retna Raj. Proposed an approach to improve the searching time of duplicated data. Dedicated hash tables were designed, each of which is used for each digital data type. When a file is received, its hash key is compared with the hash table corresponding to its type. Thus, the time required for the matching process is reduced compared to previous techniques. However, this technique did not give importance to the type of hash algorithm [15].

9- In 2024, N. A. Jaafar, this study examines the proposed model that includes computation of similarity using cosine coefficients, Euclidean similarity, and Jaccard similarity between training and test texts, providing a variety of metrics for comparison and analysis. These sequential steps combine automated analysis with human interpretation, enhancing the electiveness and accuracy of the plagiarism checker and making it easier to

with small amount of data. However, this technique is considered the least effective because the final user's data may match files on the server [8].

2- In 2016, V. Radia and D. Dingh, made a study of data deduplication techniques: file level, block level, inline post process, source based and target based. The study concluded that source-based deduplication technique is the best as it optimizes the uploading bandwidth and storage space over cloud. Distributed deduplication provides security. Both approaches together provide reliability [9].

3- In 2016, Parth Shah *et al*, proposed a technique to detect duplication between files of users. The technique involves detecting duplication not only within a user's files but also across files from different users. Once the unique files have been identified. This technique is considered more effective in saving storage space and has a medium time complexity since the process takes place at the level of users' data and not at the server or client. However, user files coming into the storage system may match files that already exist [10].

4- In 2017, Ishita Vaidya and Prof. Rajender Nath, proposed a technique to generate a hash key for the file using MD5 algorithm. Then, this key is compared with the stored keys in Hadoop [11].

5- In 2018, Manjunath R. Hudagi and Sachin A. Urabinahatti, proposed a technique to deduplicate data on file-level. This technique is based on building a hash table in the Hbase that contains the hash keys for the files stored within the system. To process each incoming file, a hash value is generated using a specific algorithm. This hash value is then compared with the values stored in the hash table. If there is no match between the generated hash and any of the stored values, the file is considered unique and subsequently stored in the system [12].

compression. The Most Common Audio Formats with Lossless Compression are:

1- Free Lossless Audio Codec (FLAC).

2- Apple Lossless Audio Codec (ALAC).

The figure (1) shows the most common audio files formats.

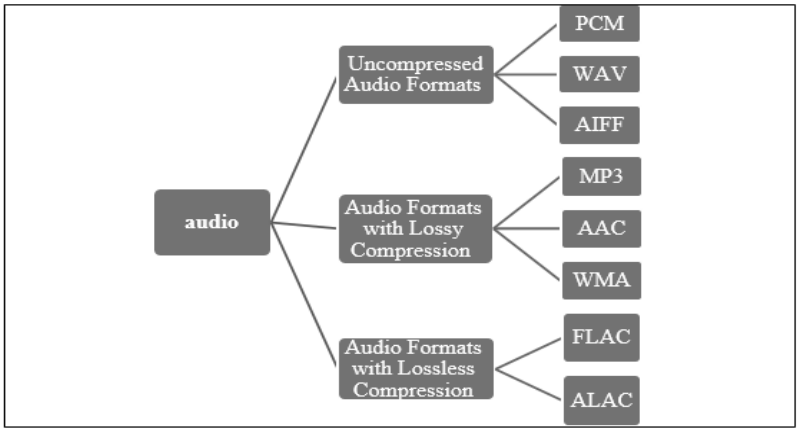


Figure (1): Most common audio files formats.

3. Related Work

In the cloud, File-Level Deduplication generally depends on generating a hash value for the incoming file using a hash algorithm and comparing it with the hash values already stored in the cloud. If this value doesn't exist in the hash table, the file is stored in the cloud and its hash value in the hash table.

We'll show some of the cloud deduplication techniques:

1- In 2015, Naveen A N and V Ravi, proposed a technique to detect duplicate user's files, and then the unique data is stored in the server. This technique is characterized by a low time-complexity, since the process works

files tend to be the most accurate and true to the original sound, providing high-quality audio reproduction, they also typically require a significant amount of disk space to store due to their large file sizes. The Most Common Uncompressed Audio Formats are:

1- Pulse-Code Modulation (PCM).
2- Waveform Audio File Format (WAV).
3- Audio Interchange File Format (AIFF).

2.2 Audio Formats with Lossy Compression

Lossy compression is characterized by some data loss that occurs during the compression process, meaning that not all original information is preserved. Despite this, compression is essential because uncompressed audio files tend to occupy a significant amount of disk space, making storage and transfer more challenging. Therefore, lossy compression helps reduce file sizes, making it more practical to store and share audio files efficiently. The Most Common Audio Formats with Lossy Compression are:

1- MPEG-1 Audio Layer 3 (MP3).
2- Advanced Audio Coding (AAC).
3- Windows Media Audio (WMA).

2.3 Audio Formats with Lossless Compression

Opposite to lossy compression, there is lossless compression, which is a technique that reduces the size of an audio file without any data loss occurring during the process. This means that the compressed audio file remains an exact replica of the original source, with all the original data preserved perfectly, ensuring no quality is compromised even after

Additionally, employing multiple indexes helps address collision problems—situations where different data produces identical hash values—by distributing hash keys across various tables, thus lowering the probability of collisions and improving the accuracy of duplicate detection [4] [18]. The selection of an appropriate hash algorithm is another critical factor influencing deduplication efficiency. The hash function's length and complexity directly affect collision rates; shorter or less robust algorithms tend to produce more collisions, increasing the risk of false positives or negatives. Therefore, choosing a cryptographic hash function with a suitable balance of speed and collision resistance is essential for effective deduplication processes [5]. Based on these considerations, we propose an algorithm that involves creating a hash table specifically for audio files on the server. This table consists of multiple indexes, each containing hash keys generated using the MD6 algorithm. MD6 is a cryptographic hash function that employs a Merkle tree-like structure, enabling efficient parallel computation of hashes for extremely long inputs [6].

2. Audio File Formats

There are three main types of audio files formats: Uncompressed Audio Formats, Audio Formats with Lossy Compression, Audio Formats with Lossless Compression [7].

2.1 Uncompressed Audio Formats

Uncompressed audio formats (UAF) consist of real sound waves that are captured and converted directly into a digital format without any additional compression or processing. Although these uncompressed audio

1. Introduction

The cloud storage system is responsible for managing and storing vast amounts of data generated from diverse sources, encompassing various formats such as audio, text, video, and images. Due to the nature of data collection and sharing, there is often a significant presence of duplicate data within the system, which leads to unnecessary consumption of storage resources and can hinder overall system performance [1]. To address these challenges, techniques aimed at eliminating duplicate data—commonly known as data deduplication—have been implemented to optimize storage utilization and enhance processing efficiency. There are two types of duplicate data detection methods: source-based detection (client-side) and target-based detection (server-side) [2]. The core process of deduplication typically involves generating a unique hash key for each file or data block, which serves as a digital fingerprint representing the content. This hash key is then compared against existing hash keys stored in a dedicated data structure, such as a hash table, to determine whether the data already exists within the system. If a matching hash is found, the system recognizes the data as a duplicate and avoids storing a redundant copy, thereby conserving storage space [3]. As the volume of stored data grows, the number of hash keys stored in the hash table also increases, which can lead to longer search times during comparison operations. This escalation in search duration can negatively impact system performance, especially in large-scale cloud environments. To mitigate such issues, one effective strategy involves creating multiple hash tables or indexes. These multiple structures can partition the stored data, enabling faster lookup operations by narrowing down search scopes and reducing the number of comparisons needed.

Abstract

Data duplication is a significant challenge in large-scale data storage systems, as it consumes storage space and impacts data organization, management, and processing. An optimal storage system effectively utilizes available storage space. To solve this problem, hash algorithms are employed to generate hash keys for files. Matching files have the same hash key. However, the hash key for two different files in the data may match, and this is what we refer to as a collision. The collision issue is related to the length of the hash key. As the length of the hash key increases, the probability of a collision occurring decreases. When a file is uploaded to the cloud storage system, its hash key is compared with the existing keys stored in the system. However, as the amount of data stored in the cloud increases, the time required for searching and matching also increases. In this paper, we will introduce a File-Level Deduplication technique to deduplicate audio data in the cloud storage system. The proposed technique aims to reduce the search time for hash values by creating a table with multiple indexes. These indexes are categorized based on the format of the audio file, such as uncompressed formats, formats with lossy compression, and formats with lossless compression. Each table contains multiple indexes in the hash table, specifically designed for a particular audio file format. To reduce the probability of data collision, Message Digest-6 (MD6) algorithm will be used, which generates a 512-bit hash key.

Keywords: Deduplication - Hash Table - MD6 - Audio Files - Cloud Storage.

Designing an Efficient Deduplication Algorithm for Audio Files in Cloud Storage / Original article

Prof. Dr. Eng. Ammar Zakzouk
Al-Esraa University – Baghdad – Iraq: ammar.zakzouk@esraa.edu.iq
Homs University – Homs - Syria: azakzouk@homs-univ.edu.sy

Assoc. Prof. Dr. Eng. Alaa Al Sebae
Warwick University – Coventry – United Kingdom: a.al-sebae.1@warwick.ac.uk

Ph.D. Student Eng. Hasan Hasan
Homs University – Homs - Syria: h_hasan@homs-univ.edu.sy

Contents

(A Written Undertaking (Pledge) of Intellectual Property)

I /We hereby certify that ..... I am/(We are) the author(s) who have achieved and written the article entitled .....

I /We confirm that this article has never been published in any other journal whether locally or internationally . I /We submit this article for consideration for publication in (**Al-Esraa University College Journal for Engineering sciences**) issued by the Al-Esraa University.

Signature (s) :

Date:

(A Written Undertaking (Pledge) of Copyrights Transfer)

I / We hereby certify that I / We ..... ,am/ are the authors of the article entitled .....

I /We agree to transfer the copyright to (**Al-Esraa University College Journal for Engineering Sciences**) issued by the Al-Esraa University.

Signature(s) :

Date:

A .Scientific research in a Journal.

Authors name, year, research title, journal name, volume, issue number and page , numbers.

B. Books.

Authors name, year, title of the book, edition, publishing house and number of pages.

C. Theses and dissertations.

Authors name, year, title of thesis, address of the college and university, and number of pages.

D. Scientific research in the proceedings of a scientific conference or symposium.

Authors name, year, the paper title, the name of the conference or the scientific symposium, venue, the starting and ending pages of the paper.

The journal is highly committed to preserving the intellectual property rights of authors.

Articles are sent to the Al-Esraa University College Journal for Engineering Sciences at the following address:

Al-Esraa University – Documentation and Scientific Publishing Department

Baghdad – Iraq

E_mail : al-esraajournal@esraa.edu.iq

- The reviewer should clearly indicate one of the three options as follows:
  - The research is suitable for publication without modifications.
  - The research is suitable for publication after changes are made.
  - The research is not suitable for publication
- The reviewer should clarify in a separate sheet the basic modifications suggested before accepting the article for publication.
- The reviewer has the right to get the manuscript back to him after making the necessary modifications to make of sure of the authors commitment.
- The reviewer must register his / her name, scientific degree work , address and the evaluation date, with the signature of the evaluation form sent, accompanied by the article submitted for evaluation.

References

1. References in the text of the manuscript are indicated as follows: The title or last name of the author and the year if the work is done by one scholar. if there were two authors they should be mentioned along with the year. In case of being three and more, the first one is mentioned then et al., and the year.
2. Reference should be listed according to (APA) and as the examples mentioned:

of Coarse Aggregate by Junk Rubber."Al-Esraa Univer. College J., 1(1), 217-243.

e.g. Garrick, G.M., (2005) " Analysis and Testing of Waste Tire Fier Modified Concerete ", M.Sc. Thesis, University of Luisiana State, U.S.A., Louisiana, pp. 9-15.

10- The abstract in English must be clear and describe the research and the results in a precise manner and not necessarily be a literal translation of the Arabic abstract and followed by 4-6 keywords.

Reviewer Guidelines

Below are the terms and requirements to be taken in consideration by the reviewer of the research sent for publication in this journal:

- Filling the evaluation form sent with the research to be evaluated accurately and not leave any paragraph without an answer.
- The reviewer must make sure that the titles, both Arabic and English, are linguistically identical. If not, an alternative title is to be suggested.
- The reviewer should state whether tables and figures seen in the research are thorough and expressive.
- The reviewer should state whether or not the authors uses statistical methods correctly.
- The reviewer should state whether the discussion of the results logically sufficient.
- The reviewer should determine the extent to which the authors uses modern scientific evidences.

2- The title of the research should be brief and expressive

3- Authors names: the names of authors and their work place addresses should be clearly written along with the first authors e-mail address.

4- An abstract should be clear and about 250- 300 words, followed by a keyword (4-6) in Arabic if the article is in Arabic language followed by abstract and keywords in English language and virus visa.

5- Introduction: includes a review of information relevant to the subject of research in the scientific sources, ending with the aim of the study and its rationale.

6- Materials and Methods : Should be fully detailed if they are new. In case of being already published, they should be mentioned in brief with reference to the sources and the use of System International Units (S.I.U.s) for measuring weight and volume.

7- Results and Discussion: should be shown in a concise, meaningful and sequential manner. The results are presented in the best form. After being referred in the results, tables and figures should be place in their designated positions.

8- The Arabic numerical system should be used in the researches submitted for publication. The discussion of the results represents a brief expression of the results and their interpretations.

9- Writing the references in the list shall include the name (s) of the authors, the publication year, the title of the research, the name of the journal, volume number, issue number and the number of pages.
e.g. Al-Khafaji, J.M., Hameed, M.H. and Kareem, H.H., (2018) " Experimental Investigation on Concrete with Paetially Replacement

(Times New Roman and Simplified Arabic), while the titles in Arabic and English should be written using 14 font size. A 2-cm margin must be left from top and bottom, and 3 cm from right and left. Articles should not exceed more than 15 pages including tables, figures, and resources taking in consideration that the whole work is written on one face of A4 papers.

2. It is not advisable to publish an article by neither the editor-in-chief nor the members of the editorial board of the journal, whether it is a solo or joint work.
3. After being approved for publication, the article is to be presented in three hard copies and an electronic one. The article is submitted in the final form by being printed on a regular basis for all pages excluding the first one which has the title of the article and the names of the authors and their addresses in both Arabic and English language in addition to the e-mail of the first author English language, the CD copy of the article should be made using Microsoft word 2010.
4. Papers may be accepted in both Arabic and English language. However, English is highly preferred.

Author Guidelines

Below are the terms and requirements need to be considered by the researcher wishing to publish in this journal:

1- The research must not published in any other engineering journal and has not been completed for more than four years prior to publication.

Guidelines of Publication in the Al-Esraa University College Journal for Engineering Sciences.

The Al-Esraa University College Journal for Engineering Sciences is published annually by the Al-Esraa University in term of two issues per year.

- The journal is concerned with publishing scientific papers as follows in the Engineering Sciences:
  - Construction engineering.
  - Civil engineering.
  - Chemical engineering.
  - Computer engineering.
  - Electrical engineering.
  - Material engineering.
  - Mechanical engineering..
  - Oil engineering.
  - …… etc.

Terms of publication

1. Each manuscript must be typed using a computer in a single spaced text on one face of the A4 paper (size A4) using 12 font size type

AL Esraa

University College Journal for Engineering Sciences

A Periodical Comprehensive Refereed Scientific Journal - Issue by: AL-Esraa University, Baghdad - Iraq

ISSN: 2709 - 7145.
E-ISSN: 2790 - 7732
The number of deposit at books and documents house,(2445), Baghdad,Iraq (2020).

Vol.(7), No.(11)-2025

ISSN: (2709 – 7145).
E-ISSN: (2790 – 7732)
The number of deposit at books and documents house, (2445), Baghdad, Iraq(2020).

Republic of Iraq
Ministry of Higher Education and Scientific Research
AL-Esraa University

AL-ESRAA University College Journal for ENGINEERING SCIENCES

A Periodical Comprehensive Refereed Scientific Journal
Issue by AL-Esraa University - Baghdad / Iraq

Volume (7) - Number (11), 2025