Approaches to Critical Digital Literacy in the MENA Region

Safae HAJJOUT, Othmane AZOUBI, Khaoula JELLAL and Mounia BADDOU
Supervised By: Dr. Loubna MEKOUAR

University Mohammed VI Polytechnic, College of Computing
Benguerir, Morocco

safae.hajjout@um6p.ma
othmane.azoubi@um6p.ma
khaoula.jellal@um6p.ma
mounia.baddou@um6p.ma
loubna.mekouar@um6p.ma

Abstract. This research paper aims to propose a viable solution to one of the world's most dire issues as of this decade, and most specifically within the MENA region. It is the proliferation of fake news, and misinformation that runs rampant, whether in community spaces or on social media. This advent has been further exacerbated by the recent rise of internet availability and content consumption in the region. Especially now that content creation, powered by automation, AI text generation and image manipulation, has never been higher. This study will focus on the most vulnerable parties of this issue: the youngest and the eldest of society, who are disproportionately susceptible. The objective of this research is to help support these vulnerable groups in developing stronger critical thinking skills so as to safeguard them against fake news in a beginning initiative, and lastly follow up with methods to prevent their spread in the first place with ethical methods. The approaches chosen to undertake this challenge will focus mainly on two facets: Firstly, case studies involving users within the MENA region clusters, both online and offline. Secondly, an analysis of how self-proclaimed news outlets manage to reach wider audiences.

Keywords: Fake news. Misinformations. Media Bias. AI. Social media. Digital literacy. Critical thinking. Education.

Introduction

The primary focus of this research paper is an analysis of the intricate and multifaceted factors contributing to the spread of misinformation and fake news within the MENA region, with the overarching objective of developing effective counter methods both in the social and technological dimensions.

The framework of this analysis was helped by the incorporation of a comprehensive problem examination, as the core question: "How do we help develop digital literacy in the MENA region and gain back trust in reliable news sources?" entails, achieved by case studies involving MENA citizens across the following clusters: age group, education level and region. A thorough investigation split across two primary timelines: the aftermath of the Arab Spring, and post-Covid19, we aim to understand the local citizenry's perceptions and how they interact with information sources such as media outlets, online social media posts, and national news coverage. In sections 3.1 and 3.2 we will give shape to the intellectual landscape of the MENA region, especially on digital platforms.

These case studies, as we will later see within this paper, were fruitful in helping us explore our options and compose them into strategies that will help counter the spread of fake news and promote digital literacy within user groups, and more generally develop their critical thinking skills, especially important in the context of the expanding and increasingly complex and hostile digital environment, compounded by the advent of generative AI that only further complicates the task of finding what is true and what is not. The implementation of these endeavors will be further discussed in section 4.1.

In section 4.2, our proposed algorithm, aided by a thorough study of the Credulix plugin [1], will shed light into how we can prevent the spread of fake news in the first place, before they reach wider audiences.

2. Previous work

The algorithmic aspect of this research builds upon the pivotal studies that have significantly contributed to the domain of fake news detection. A particularly noteworthy contribution is that of Pr. Alexandre Maurer and his colleagues, indeed in the article "Limiting the Spread of Fake News on Social Media Platforms by Evaluating Users' Trustworthiness." The plugin, Credulix, is introduced as a plausible solution. It is a content-agnostic system that bases itself upon users' reputation to determine the nature of their posts, whether they are accurate news or falsified information. Its process is as follows: Credulix analyzes a user's previous sharing behavior to estimate the likelihood of their unchecked news items being misinformed and thus flagging them. This reputation based system as is, works well in theory and controlled environments with the proliferation of misinformation while maintaining minimal impact on overall application performance. However, the practical application of Credulix and its broader impact in diverse contexts, such as the MENA region, presents challenges that warrant further improvement. Firstly, Credulix is not a standalone solution but a complementary tool that augments manual fact-checking efforts. It relies on a fact-checking team to provide a foundational ground truth. Secondly, the premise of Credulix's methodology assumes consistency in user behavior—a hypothesis that may not hold against the backdrop of evolving misinformation campaigns along with generative AI.

In recognizing the limitations of existing models, our study seeks to advance the field by exploring methodologies that can dynamically adapt to changing user behaviors and emerging tactics employed by purveyors of fake news.

3. Field Studies

In the following segment, we delve into the studies done on the people of some MENA countries,i.e: Egypt, Morocco, Qatar, UAE and others followed by some commentaries on these statistics. We finally show some conclusions that will be crucial to the development of our solutions in relation to our initial problem.

These figures represent two pivotal timelines for the MENA region:
Post Arab Spring (2015-2019) and Post Covid-19 (2020 -).

3.1 Post Arab Spring MENA:

The Arab Spring, a series of pro-democracy uprisings that unfolded across the Arab world in the early 21st century, represents a transformative and complex socio-political phenomenon. Emerging in late 2010, the movement was characterized by widespread popular discontent with authoritarian regimes, socioeconomic grievances, and demands for political reform. The wave of protests quickly spread to various Arab countries, including Egypt, Libya, Yemen, and Syria.

The aftermath of these uprisings witnessed a notable decline in trust toward traditional news outlets among affected populations. The demonstrations, fueled by widespread dissatisfaction with authoritarian regimes, coincided with the rise of social media platforms. During this period, citizens increasingly turned to platforms like Twitter, Facebook and YouTube as alternative sources of information, drawn to the immediacy and perceived authenticity of user-generated content.

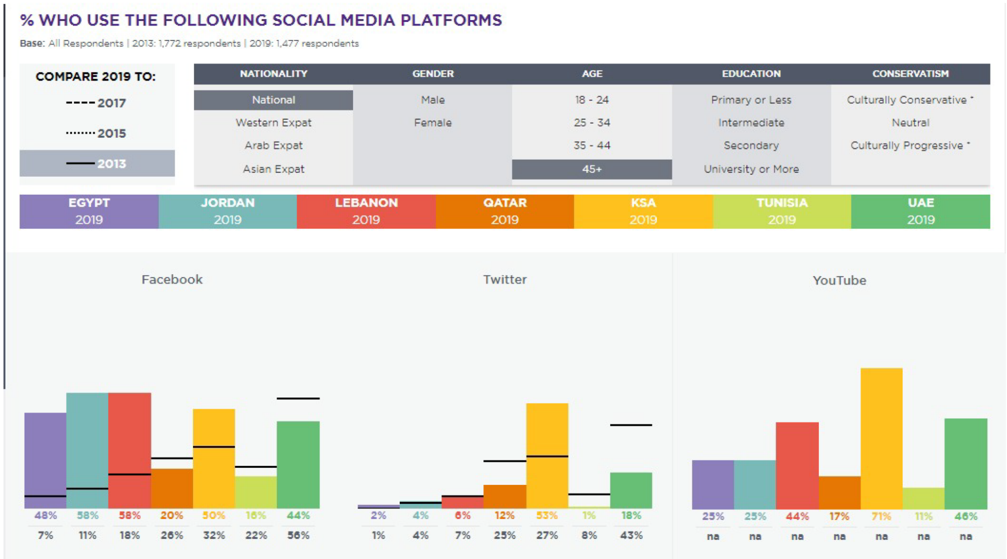

Figure 1: Various Stats about Facebook, Twitter and YouTube across MENA region for elders between 2013 and 2019

The examination of social media platform usage among the elderly demographic (45+ years) in various MENA countries during the period from 2013 to 2019 is presented in Figure 1 [2], revealing a distinct contrast. Notably, Facebook emerges as the predominant platform, experiencing increased adoption among older individuals in countries such as Egypt, Lebanon, and KSA. Conversely, Twitter registers comparatively lower usage among the elderly, with the exception of KSA where a notable doubling in adoption is observed. These findings suggest potential correlations with factors such as user interface friendliness or the perceived quality of information derived from each platform.

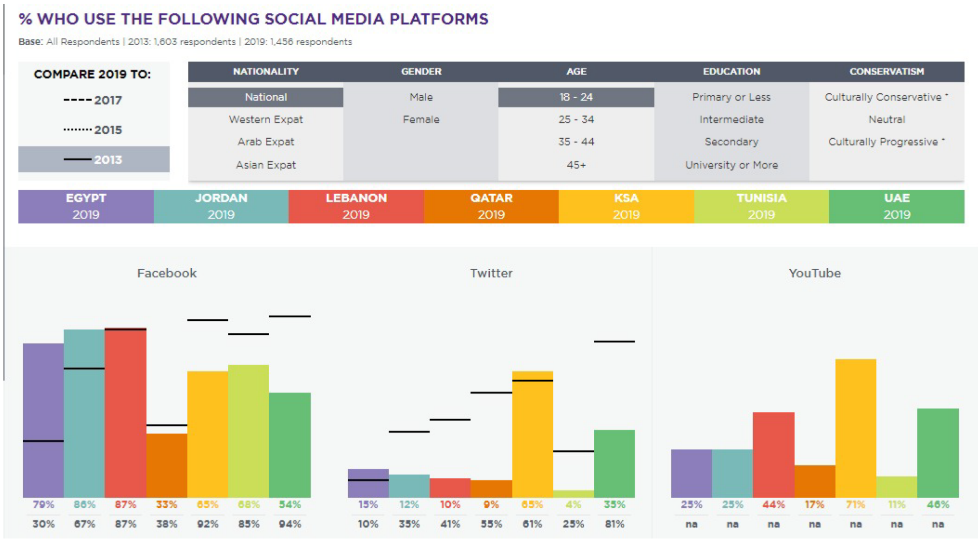

Figure 2: Various Stats about Facebook, Twitter and YouTube across MENA region for the younger generation between 2013 and 2019

In Figure 2 [2], we turn our attention to the usage trends of social media platforms among the younger demographic (18-24 years) across various MENA countries from 2013 to 2019. Striking divergences in platform adoption become apparent, with increases noted in countries such as Lebanon and Egypt, and conversely, declines observed in Qatar, KSA, and UAE. These fluctuations may be attributed to several factors, including shifts in platform popularity and the emergence of new alternatives like WhatsApp and TikTok. Additionally, the dynamic landscape may reflect growing dissatisfaction among the youth regarding the content disseminated through traditional platforms.

3.2 Post Covid-19 MENA:

The COVID-19 lockdowns in the MENA region brought about a significant change in how people interacted with digital information [3]. As individuals spent more time at home, their reliance on digital devices for information increased, leading to a noticeable shift in attitudes towards online content consumption. The rapidly changing nature of the pandemic, combined with the initial lack of clear guidance, led to a heightened dependence on digital platforms as primary sources of information. People turned to these platforms to learn about new developments, ways to stay safe, and potential treatments.

However, this surge in digital information consumption came with a downside. Misinformation and unverified claims spread widely during this period, creating challenges [4]. The uncertainty surrounding the emerging health crisis and the scarcity of reliable information made individuals more vulnerable to unchecked content from online sources. Consequently, this vulnerability contributed to issues such as widespread doubt about the seriousness of the pandemic and the unchecked adoption of unverified remedies promoted by online communities. This complex interplay underscores the challenges associated with increased reliance on digital information, particularly during times of uncertainty like the COVID-19 lockdowns in the MENA region.

4. Solutions

As previously indicated, we will structure our solutions by beginning with the proposed social approach, according to age groups, then the algorithmic solution further explained below.

4.1. Social Aspect

4.1.1) Credible News in Social Media

As the field studies elucidated, and keeping in mind that the youth are the most prevalent age group across the MENA region, it is imperative to consider how to reach them first and foremost.

Our first insight was which social media are predominantly used. We found that they were, by order: Whatsapp, Youtube then Facebook and Instagram, for most countries. While the statistics vary, the case studies prove these are the most consistently used. Thus, credible news media sources' firstmost goal should be to appeal to these audiences and see how they might reach them. It is equally suggested [5] that quick format news, by which we allude to simple pictures with simple, short and clear text accompanying them, is the most consumed type of content. This simple transformation in how news are formatted can assure the reach to wider audiences. In a side note, this quick format should not be abused, as it equally favoritizes transforming the journalism sector into another quick content consumption method over what it essentially presents: a sector that requires careful thought and ideas assessment while navigating biases. Interactive and enticing graphics should not be the main focus, but rather the content they attempt to shed into light. Ted exemplifies a better way of handling news online, as their videos are short and to the point yet remarkably informative and thought provoking, including their talks [6].

4.1.2) Critical Thinking in Education

Critical thinking is a hard learned skill that is not as innate as some may think [7]. Yet despite its importance within the digital literacy context, emphasis is rarely put on it in education. As Alsaleh suggests in her paper "Teaching Critical Thinking Skills: Literature Review" [7], the integration of critical thinking within school curriculums across all grade levels is crucial. As for the strategy in implementing it, two approaches are suggested. One deems it more favorable to include critical thinking as a separate module to be learned, the same way as mathematics, writing and sciences are taught. The other favors instead its implicit inclusion within each module. This latter method, while undoubtedly more efficient according to research, will require more work on the educators' side in creating environments that place critical thinking at the forefront while also maintaining engagement.

4.1.3) Digital literacy in Education

Complementary to this point, digital education within the MENA region, according to our field studies, is lacking. Despite internet users' percentages ever growing, going from 60% to 99% [8] to population, in the MENA region, most do not know how to properly navigate the digital landscape of the Internet. Addressing this gap is essential, doing so resides in designing digital programs, accompanied by institutes such as our own, Mohammed VI Polytechnic University in Morocco, and others. The main objectives should be teaching essential skills such as knowing how to search for and look up information, recognizing the dangers of the internet, from scams to our focus, the sharing of fake news. These programs; preferably to be within the scope of official school curriculums to reach as many as possible; in combination with the previously discussed critical thinking implementations will undoubtedly yield a generation killed in questioning assumptions, analyzing evidence, and making informed judgments.

4.1.4) Approaching Older Generations

While the suggested approaches cited above work well for the younger generations, they cannot be applied as effectively on older generations. This, however, is not a shortcoming. In our research, we have thought of a few critical ways by which to approach this problem. Starting from the basis that by influencing the youth, they may equally influence their entourage and community including elders. This is a vital idea to keep in mind. Additionally, the phenomenon of social media influencers could be exploited within our favor by forming partnerships, as they tend to have wider reach than official programs, their popularity could be in our favor. We may illustrate this by two examples in North Africa that come to mind: the acclaimed youtuber "Swinga" in Morocco is well known for his educational content, as well as "AlBernameg" [9] in Egypt, a television program that presents important facts; for instance during the Arab Spring; and questions unfounded misinformation through the medium of comedy and satire. These influencers do great work in favor of critical thinking and deterring falsified news, thus by partnering with institutes they may change the way information is processed in the MENA region via launching campaigns, publishing more content, audiovisual and otherwise.

4.2. Algorithmic Aspect

4.2.1) Definitions:

Based on “The Fake News Vaccine” research article [1], we extract the following definitions:
Let $g \in (0,1)$ be the global fraction of fake news in the desired social media.
Let U be the a user with the following attributes:

$vT(u)$: Number of fact-checked true items viewed by u
$sT(u)$: Number of fact-checked true items shared by u
$vF(u)$: Number of fact-checked false items viewed by u
$sF(u)$: Number of fact-checked false items shared by u

The tuple $(vT(u), sT(u), vF(u), sF(u))$ is called the User Credulity Record (UCR) of u.

Let X be a news item, the research article defines:

$$\beta_1(u) = \frac{sT(u)+1}{vT(u)+2}$$

$$\beta_2(u) = \frac{sF(u)+1}{vF(u)+2}$$

$$\beta_3(u) = \frac{vT(u)-sT(u)+1}{vT(u)+2}$$

$$\beta_4(u) = \frac{vF(u)-sF(u)+1}{vF(u)+2}$$

$$\pi_T(V,S) = \prod_{u\in S}\beta_1(u)\cdot\prod_{u\in V-S}\beta_3(u)$$

$$\pi_F(V,S) = \prod_{u\in S}\beta_2(u)\cdot\prod_{u\in V-S}\beta_4(u)$$

Let V (resp. S) represent the set of users who have viewed (resp. shared) item X.
Let p_0 be the threshold probability of an item being fake; it is a constant defined in the system.

The formula for the item rating $\alpha(V,S)$ is defined as follows:

$$\alpha(V,S) = \frac{\pi_T(V,S)}{\pi_F(V,S)}$$

The rating threshold α_0 corresponding to a probability threshold p_0 is defined as:

$$\alpha_0 = \frac{1}{p_0} - \frac{1}{g-1}$$

An item X is considered fake when: $\alpha(V,S) \leq \alpha_0$.

4.2.2) Automated Ground Truth Evolution:

For Post-Based social media:
In the MENA region, where platforms like YouTube, Facebook, and X (previously known as Twitter) social media are predominant, our approach involves a two-pronged process for post-based social media.

1. Extraction and Analysis Process:
The initial stage involves extracting key elements from various social media platforms. This includes:

Content Descriptions and Titles from YouTube, Facebook, and X social media:
We collect detailed descriptions and titles associated with the media.

Media Processing Algorithm:
Concurrently, we focus on the 'most watched’ or highly engaged sections of other platforms like Facebook and X social media, targeting content that garners significant user attention.

Following the text extraction (descriptions and titles) phase, we deploy a media processing or extraction algorithm; of which many currently exist for analytic studies and content recommendation like Matrix Factorization [10] and clustering algorithms like K-means [11]; tailored to analyze the 'most viewed' sections of these platforms and check their context in comparison to the given title and description. To summarize, the algorithm sifts through the most popular segments of the platforms, capturing the essence of what is most viewed or interacted with by users and finally comparing it to the textual context.

2. NLP-Based Comparison and Verification:

Applying NLP techniques:
At this juncture, advanced Natural Language Processing (NLP) techniques, (i.e Semantic Analysis Algorithm [12]) are employed to compare the extracted descriptions and content from media with the data extracted from the most viewed content.

Difference Percentage (d):
The outcome of this comparison is quantified as a percentage difference, denoted as 'd'. This metric encapsulates the extent of disparity between the presented media and its corresponding description or title.

Fake news identification:
Threshold Determination d_0: A predefined threshold d_0, is established to categorize news items. If the calculated difference percentage 'd' surpasses this threshold, **the content is flagged as potentially fake**.

For Chat-Based social media:
In adapting our methodology to chat-based social media platforms, such as WhatsApp, the fundamental process remains consistent with that employed for post-based platforms like YouTube and Facebook. The core steps of extraction, analysis, and verification are mirrored, with tailored adjustments to suit the specific content types prevalent in chat-based environments. It is also of note that some of Whatsapp's features such as the "forwarded many times" caption may be used in tandem with the algorithm.

4.2.3) User suspicion record:

In our approach to patching the main vulnerabilities (assuming that users' behavior generally remains stable over short time frames) in reputation systems. We thought of using the User Suspicion Record (USR) as a key metric.

To build USR, we collect and analyze data based on multiple user behavior indicators. These include how often users share information (whether flagged or unflagged) and any unusual behavior in their account activity.

This score serves as an indicator of a user's likelihood of trying to exploit the system's assumption to spread misinformation. It's important to note that USR evolves in response to the user's ongoing activities.

We define:
$Us(u)$: The percentage of suspicion in a user u
S_0 : The suspicion threshold for a user to be reported to the moderation team
A user u is reported to the moderation team when: $Us(u) \leq S_0$

When a user is reported to the moderation team, they get to change his credulity record based on their analysis.

4.2.4) Fighting malicious use:

The assumption that users won't quickly change their behavior can be exploited by a malicious actor "X," employing numerous machine-operated users (bots). The seemingly straightforward solution is to monitor the items shared by "X" until they become viral and then fact-check them. However, once an item becomes viral, it may be too late, and many people might have already seen it. To address this challenge, we propose:

1. Dealing with Malicious Actors Using Bots:
- To counteract the influence of machine-operated users, we propose the introduction of multilayer authentication and captcha systems, this measure targets items shared by users identified as potential botnets or similar automated entities.
- The goal is to limit automated account activity, uphold the system's assumption about consistent user behavior and prevent rapid, unverified information spread.

2. Handling Collaborative Malicious Actors:
- To address the issue of collaborative malicious actors, we recommend employing diverse content analysis algorithms as those mentioned in section 4.2.2.
- These algorithms function by independently evaluating each actor's behavior and subsequently updating their user suspicion value $Us(u)$.
- By analyzing user-generated content, such as posted thumbnails, usernames and profile pictures, we establish patterns to identify users with high suspicion values.
- This method enhances the system's capacity to detect and address efforts in spreading misinformation.

Conclusion

As we conclude our paper, spanning from the foundational inquiries delineation to the methodology pursued to reach our final research, many new insights were gained. Firstly we began by laying out our groundwork on case studies in the MENA region following the Arab Spring and post Covid-19 pandemic timelines. Subsequently, we applied solutions based on our findings and local population studies, that mainly targeted in an initial social dimension the digital literacy aspect and general implementation of critical thinking. We finally focused our efforts on the algorithmic side, built upon previous research (the Credulix plugin, content processing algorithms etc.) in order to effectively counter fake news spread.

References

1. *Balmau, O., Guerraoui, R., Kermarrec, A., Maurer, A., Pavlovič, M., & Zwaenepoel, W. (2018). Limiting the spread of fake news on social media platforms by evaluating users' trustworthiness. arXiv (Cornell University). http://export.arxiv.org/pdf/1808.09922*

2. *Interactive · Media Use in the Middle East, 2019. (n.d.). https://www.mideastmedia.org/survey/2019/interactive/social-media/who-use-the-following-platforms.html*

3. *Jahan, I., Hosen, I., Al‐Mamun, F., Kaggwa, M. M., Griffiths, M. D., & Mamun, M. A. (2021). How has the COVID-19 pandemic impacted internet use behaviors and facilitated problematic internet use? a Bangladeshi study. Psychology Research and Behavior Management, Volume 14, 1127–1138. https://doi.org/10.2147/prbm.s323570*

4. *Nelson, T. B., Kagan, N., Critchlow, C., Hillard, A. E., & Hsu, A. (2020). The danger of misinformation in the COVID-19 crisis. Missouri Medicine, 117(6), 510–512. https://pubmed.ncbi.nlm.nih.gov/33311767/*

5. *Nhedzi, A. (2019). The media of Consumption and the Consumption of time: How a consumer in fast-paced economy use traditional and new media tools. Observatorio (OBS*), 13(2). https://doi.org/10.15847/obsobs13220191345*

6. *https://www.youtube.com/@TED/videos*

7. *Alsaleh, N. J. (2020). Teaching Critical Thinking Skills: literature review. Turkish Online Journal of Educational Technology, 19(1), 21–39. http://files.eric.ed.gov/fulltext/EJ1239945.pdf*

8. *DataReportal – Global Digital Insights. (n.d.). DataReportal – Global Digital Insights. https://datareportal.com/search?q=mena*

9. *Bassem Youssef's El-Bernameg gone for good - Politics - Egypt. (n.d.). Ahram Online. https://english.ahram.org.eg/News/102760.aspx*

10. *Koren, Y., Bell, R. M., & Volinsky, C. (2009). Matrix factorization techniques for recommender Systems. IEEE Computer, 42(8), 30–37. https://doi.org/10.1109/mc.2009.263*

11. *Hartigan, J., & Wong, M. A. (1979). Algorithm AS 136: A K-Means clustering algorithm. Journal of the Royal Statistical Society Series C: Applied Statistics, 28(1), 100. https://doi.org/10.2307/2346830*

12. *Bellegarda, J. R., Butzberger, J. W., Chow, Y. L., Coccaro, N. B., & Naik, D. (1996, May). A novel word clustering algorithm based on latent semantic analysis. In 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings (Vol. 1, pp. 172-175). IEEE.*

Low Power, High Impact: LoRaWAN for Secure Wearable Health Monitoring in Energy-Scarce Regions

Othmane Azoubi
College of Computing
Mohammed VI Polytechnic University
Benguerir, Morocco
othmane.azoubi@um6p.ma

Khaoula Jellal
College of Computing
Mohammed VI Polytechnic University
Benguerir, Morocco
khaoula.jellal@um6p.ma

Loubna Mekouar
College of Computing
Mohammed VI Polytechnic University
Benguerir, Morocco
loubna.mekouar@um6p.ma

***Abstract*—Throughout the years, the challenge of maintaining energy in environments where reliable power sources are scarce or unavailable has posed a constant issue, especially for wearable health monitoring devices that need to operate continuously to guarantee patient safety and facilitate effective healthcare delivery. With the rising demand for these devices, numerous innovative solutions have been developed.**

This paper discusses the feasibility of integrating low-energy communication protocols into wearable devices as a practical strategy to minimize energy consumption while enhancing communication range. In particular, LoRaWAN stands out as a promising choice due to its ability to provide long-range, low-power communication. This makes it especially suitable for health monitoring applications in energy-limited environments, where reliable, energy-efficient communication is essential for the continuous functionality of wearable devices.

Although LoRaWAN offers an energy-efficient communication framework for IoT networks, significant security and privacy vulnerabilities raise serious threats regarding data availability, authentication, and privacy in health monitoring applications. This paper presents a holistic framework to improve the security of health data transmission over LoRaWAN by analyzing its architecture and providing strategies to mitigate its vulnerabilities.

***Index Terms*—LoRaWAN, wearable devices, healthcare, IoT, security, energy efficiency**

I. Introduction

The primary focus of this research paper is an exploration of the transformative impact of wearable technology on healthcare, specifically in energy-challenged environments where uninterrupted health monitoring is vital. Wearable devices, which have evolved from basic fitness trackers to complex medical gadgets, now play a pivotal role in real-time health monitoring. However, this technological leap comes with pressing concerns regarding the security, privacy, and reliability of sensitive health data transmission, particularly in scenarios with constrained energy resources and limited connectivity.

The framework of this analysis is built upon addressing a central question: "How can low-energy communication protocols be optimized to ensure both efficient performance and uncompromising security in wearable health devices, especially in energy-challenged environments?" This inquiry requires a thorough examination of contexts such as rural healthcare settings and natural disaster zones, where energy constraints and security vulnerabilities converge. Conventional communication methods like Wi-Fi, Bluetooth, and cellular networks, though widely adopted, are energy-intensive and unsuitable for wearable healthcare devices operating in such settings.

Among emerging low-energy communication protocols, LoRaWAN has gained significant attention for its ability to facilitate long-range, low-power data transmission. This capability makes it uniquely suited to wearable healthcare applications in remote or resource-constrained environments. However, despite its promise, LoRaWAN faces limitations, particularly in safeguarding highly sensitive health data from potential breaches. Through a meticulous analysis of LoRaWAN's architecture and functionalities, this research seeks to establish a robust understanding of its potential and limitations in healthcare applications, with the overarching objective of proposing a secure and efficient framework for its integration.

As detailed in sections 3.2 and 3.3, this investigation contextualizes LoRaWAN's implementation within real-world use cases, critically examining its architecture and comparative advantages in energy-challenged scenarios. The analysis will culminate in a strategic framework aimed at addressing key security and privacy concerns, as outlined in section 6. Our proposed solution, informed by a comprehensive review of existing security measures, aspires to mitigate risks and enhance trust in wearable health technologies. This framework includes a novel algorithm designed to bolster the protocol's resilience, ensuring the integrity and privacy of patient data even in the most challenging conditions.

II. Background and Related Work

The foundation of this research lies in the intersection of IoT technologies and wearable devices, which have revolutionized healthcare through real-time monitoring and personalized care. A pivotal study by Abdulmalek et al. [1] introduces a hybrid healthcare system combining Wi-Fi and LoRaWAN for efficient, long-range data transmission. While accurate and reliable, LoRaWAN's inherent vulnerabilities, such as weak encryption and susceptibility to breaches, pose significant risks to safeguarding sensitive health data in practical applications. Addressing energy constraints, Kang and Yeo [2] examine energy harvesting technologies, like converting body heat or motion into power, as alternatives to traditional batteries, but their inconsistent outputs limit integration into systems requiring uninterrupted operation.

From a security standpoint, LoRaWAN faces replay attacks, eavesdropping, and key management issues. While OTAA dynamically generates session keys during join procedures for enhanced security compared to ABP, both methods struggle to ensure robust encryption for battery-constrained IoT devices [3], [4].

Emerging studies propose solutions, such as Sanchez-Iborra et al.'s lightweight key management schemes using Ephemeral Diffie-Hellman Over COSE (EDHOC) for secure, efficient session key updates [5]. Despite advancements, LoRaWAN's vulnerabilities in high-security environments like healthcare persist, necessitating frameworks that integrate robust encryption with energy efficiency.

III. LoRaWAN for Wearable Devices

A. Overview of LoRaWAN

LoRaWAN (Long Range Wide Area Network) is a low-power protocol designed for long-range communication, making it ideal for wearable healthcare devices operating in remote or energylimited settings where continuous patient monitoring is critical.

B. LoRaWAN Architecture

Unlike traditional networking protocols, LoRaWAN relies on the Aloha method [6] for data transmission. In this setup, end devices only send data when triggered by specific events, such as environmental changes detected by sensors or time-based triggers. After transmitting an uplink, these devices briefly listen for a response from the network before entering a low-power sleep state, consuming less than one microampere of energy. This approach enables wearable devices to operate on minimal power, with a battery lifespan of up to 10–15 years, ideal for long-term health monitoring in remote or resource-constrained settings.

The LoRaWAN architecture employs a star topology with three core components: end devices, gateways, and network servers. End devices, such as wearable health sensors, collect data like heart rate or blood pressure and broadcast packets asynchronously to the network. These packets are received by one or more gateways—simple, cost-effective devices equipped with multi-channel and multi-data rate radios. Gateways merely forward the data to the core network server without processing it, minimizing energy demands and enabling seamless roaming across cells without requiring synchronization.

In healthcare applications, the network server's ability to manage data using metrics like RSSI and SNR [7] ensures reliable communication for wearable devices monitoring critical vitals such as heart rate and oxygen levels. Its Adaptive Data Rate (ADR) policy optimizes battery use, enabling devices to operate for years without replacement, even in resource-constrained environments. Furthermore, the ADR policy, controlled by the network server, dynamically adjusts data rates and transmission power based on link quality, ensuring energy-efficient communication.

In some configurations, application servers are integrated with the network server, while in others, they operate independently. This flexibility allows for multi-tenant network scenarios, where diverse applications, such as remote patient monitoring or emergency alerts, coexist within the same infrastructure.

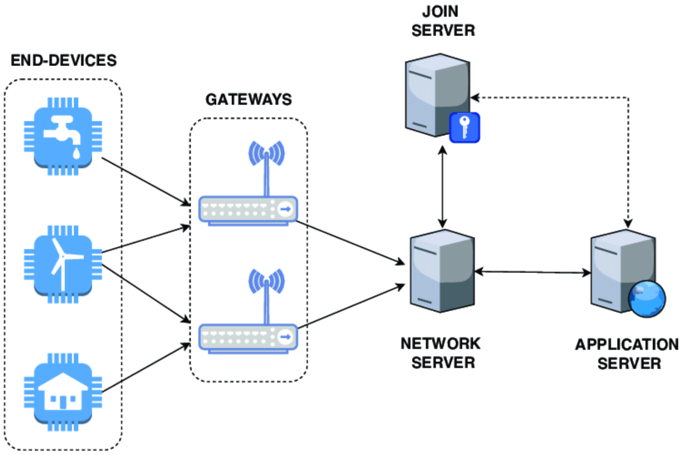

Fig. 1. LoRaWAN Architecture

C. Comparison with Other Protocols

In the context of wearable healthcare devices, choosing an appropriate communication protocol is critical, especially in energy-challenged environments where long-term functionality and minimal energy consumption are paramount. LoRaWAN stands out when compared to other commonly used protocols like Wi-Fi, Bluetooth, ZigBee, and other LPWAN technologies such as NB-IOT and Dash7 due to its unique balance of low-power consumption, long-range communication, and scalability. For instance, in rural health clinics, LoRaWAN enables continuous monitoring of patient vitals across kilometers, a task that Wi-Fi cannot handle due to its limited range and higher energy demands [1].

What sets LoRaWAN further apart is its ability to support the deployment of thousands of devices within a single network, a crucial feature for large-scale health monitoring in hospitals or rural regions. Additionally, LoRaWAN can dynamically adjust data rates and transmission power to ensure efficient energy use and reliable communication, even in the most challenging environments, thereby extending device lifespans and enhancing operational consistency.

IV. Importance of Health Data and Security

Health data is inherently more sensitive than most other types of information; it includes highly personal details such as medical histories and treatments. This sensitivity makes health data a primary target for misuse. In this context, wearable devices present multiple challenges as they generate continuous, real-time data streams, which are particularly attractive to attackers. In energy-constrained environments, such as rural clinics or disaster zones, this challenge is further amplified by limited infrastructure to support robust data protection. The continuous, real-time nature of data generated by wearable devices makes implementing traditional security protocols especially difficult when energy overhead must remain minimal.

Safeguarding health data in these settings therefore requires innovative solutions that balance strong security with minimal computational overhead. This creates a pressing need for lightweight, energy-efficient security frameworks that operate effectively under constrained conditions—ultimately reducing risks of unauthorized access, data breaches, or device malfunctions, and supporting the broader goal of accessible healthcare everywhere.

V. Security Challenges in LoRaWAN

A. Overview of the Existing LoRaWAN Security Framework

LoRaWAN incorporates various mechanisms to ensure secure communication between devices, network servers, and application servers. These mechanisms are designed to meet the requirements of the CIA model (Confidentiality, Integrity, and Availability) [4]. Key features include:

- **Key Management:** LoRaWAN uses symmetric key cryptography, where session keys are derived from a root key. Keys are shared via two activation modes: Over-the-Air Activation (OTAA) and Activation by Personalization (ABP) [8], [9].
- **Message Integrity:** LoRaWAN ensures message integrity by attaching a Message Integrity Code (MIC) to each message. The MIC is a cryptographic checksum, which helps verify that the message has not been tampered with during transmission [4].
- **Frame Counters (FCnt):** To prevent replay attacks, FCntUp (for uplinks) and FCntDown (for downlinks) are used. These counters increment with each transmission, ensuring each message is unique and preventing the reuse of old messages [4], [10].
- **Encryption:** Data confidentiality is achieved by encrypting communication between devices and servers using CTR-AES-128 encryption. The session keys (NwkSKey and AppSKey) are used for encrypting the payload [4], [9].

B. Identification of Security Vulnerabilities

Despite these security features, LoRaWAN still faces several vulnerabilities:

- **Replay Attacks:**
  - An attacker can replay a join-accept message before the Network Server (NS) delivers the authenticated one to the End Device (ED). This disrupts availability since further transmissions will be rejected [11].
  - An attacker may intercept a join-request message containing DevNonce1 and use RF jamming to block the ED from receiving the join-accept. After a timeout, the ED sends a new join-request with DevNonce2, but the attacker jams this too and replays the original join-request with DevNonce1. Since DevNonce1 is still valid, the NS accepts it, causing device and server desynchronization [10].
  - In ABP mode, static frame counters allow an attacker to capture and replay an uplink packet when the frame counter resets to 0 [12].
- **Man-in-the-Middle (MITM) Attack:**
  - An attacker intercepts data packets from the ED before they reach the gateway. After intercepting, the attacker modifies the payload and recalculates the MIC using a compromised NwkSKey. The NS then verifies this MIC and accepts the altered message, assuming it to be authentic [3].
- **Eavesdropping:**
  - LoRaWAN uses AES in Counter (CTR) mode for data confidentiality. In ABP mode, session keys are static [4], which can lead to the reuse of keys once the counter overflows [3].
  - If an attacker obtains a known plaintext P_1 and its ciphertext C_1, they can potentially decrypt other ciphertexts by leveraging XOR properties in CTR mode, thus breaking confidentiality [9].
- **Bit Flipping Attack:**
  - CTR mode uses XOR operations on each bit independently, allowing malicious alteration of bits in transit. This can corrupt data without detection if integrity checks are insufficient [3], [9].
- **ACK Spoofing Attack:**
  - An attacker captures a valid ACK from the NS and replays it to falsely confirm receipt of an uplink message from the ED. By blocking legitimate ACKs with interference and replaying old ones, the ED will wrongly believe the message was received [3], [12].

VI. A Framework for Enhancing Security with Low Computational and Energy Overhead

A. Framework Overview

Many of the vulnerabilities in Section V-B arise from static keys, Nonces, or counters. To address these issues, the End Device (ED) and Network Server (NS) will publicly agree on an initial seed, λ. They then perform a key exchange using EDHOC (Ephemeral Diffie-Hellman Over COSE), optimized for lightweight IoT [13]. EDHOC derives the initial session keys (NwkSKey and AppSKey), eliminating reliance on a root key.

- **NwkSKey as OTP Secret:**
    - The NwkSKey acts as the One-Time Password (OTP) secret. We generate an array of values from a hash function $f^{(\lambda)}$(NwkSKey), each corresponding to a unique OTP [14]. After each transmission, λ decrements, limiting OTP reuse and forcing frequent key renewal.
- **AES for Encryption:**
    - We still use AES-128, but rely on $f^{(\lambda-k)}$(NwkSKey) to serve as a dynamic counter (nonce) for each encryption, ensuring uniqueness and security. AppSKey remains the encryption key.
- **Key Renewal via EDHOC:**
    - When λ reaches 1, EDHOC is re-run to generate fresh session keys, limiting the impact of compromised keys.
- **FCnt Replacement:**
    - Since every message includes a unique OTP, frame counters (FCntUp and FCntDown) become unnecessary. This reduces the chance of replay attacks using stale counters.

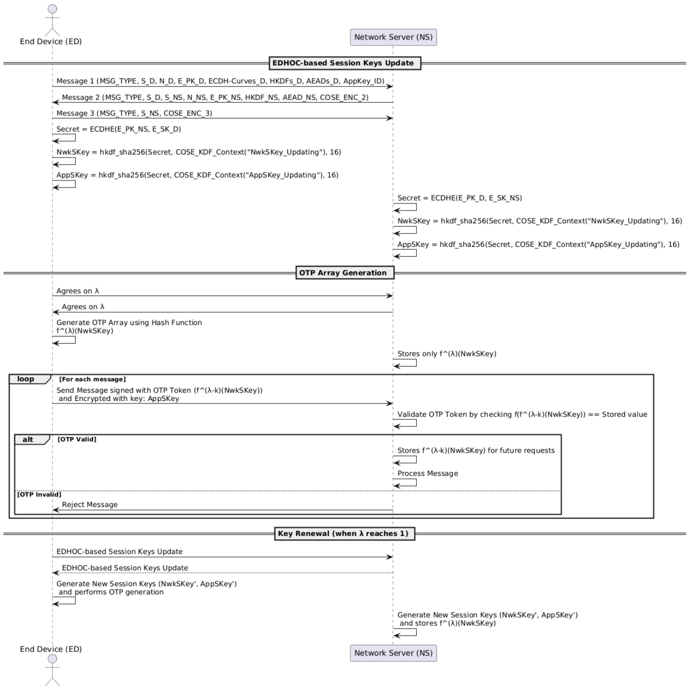
End Device (ED)
Network Server (NS)
EDHOC-based Session Keys Update
Message 1 (MSG_TYPE, S_D, N_D, E_PK_D, ECDH-Curves_D, HKDFs_D, AEADs_D, AppKey_ID)
Message 2 (MSG_TYPE, S_D, S_NS, N_NS, E_PK_NS, HKDF_NS, AEAD_NS, COSE_ENC_2)
Message 3 (MSG_TYPE, S_NS, COSE_ENC_3)
Secret = ECDHE(E_PK_NS, E_SK_D)
NwkSKey = hkdf_sha256(Secret, COSE_KDF_Context("NwkSKey_Updating"), 16)
AppSKey = hkdf_sha256(Secret, COSE_KDF_Context("AppSKey_Updating"), 16)
Secret = ECDHE(E_PK_D, E_SK_NS)
NwkSKey = hkdf_sha256(Secret, COSE_KDF_Context("NwkSKey_Updating"), 16)
AppSKey = hkdf_sha256(Secret, COSE_KDF_Context("AppSKey_Updating"), 16)
OTP Array Generation
Agrees on λ
Agrees on λ
Generate OTP Array using Hash Function
f^(λ)(NwkSKey)
Stores only f^(λ)(NwkSKey)
loop
[For each message]
Send Message signed with OTP Token (f^(λ-k)(NwkSKey))
and Encrypted with key: AppSKey
Validate OTP Token by checking f(f^(λ-k)(NwkSKey)) == Stored value
alt
[OTP Valid]
Stores f^(λ-k)(NwkSKey) for future requests
Process Message
[OTP Invalid]
Reject Message
Key Renewal (when λ reaches 1)
EDHOC-based Session Keys Update
EDHOC-based Session Keys Update
Generate New Session Keys (NwkSKey', AppSKey')
and performs OTP generation
Generate New Session Keys (NwkSKey', AppSKey')
and stores f^(λ)(NwkSKey)
End Device (ED)
Network Server (NS)

Fig. 2. Framework for enhancing LoRaWAN security

B. Vulnerabilities Countermeasures

ACK Spoofing & Replay: Each message includes a one-time token valid for a single transmission, preventing replays. Reusing a token will fail verification, causing a timeout that triggers EDHOC to resynchronize the OTP array. Eliminating ABP mode also thwarts static-key exploits.

MITM Attacks: Because NwkSKey and AppSKey refresh after λ transmissions and because EDHOC thwarts key compromises, forging a valid MIC becomes substantially more difficult.

Eavesdropping: Session keys are frequently updated, preventing exploitation of static keys over many transmissions.

Bit Flipping: The integrity verification—now checked end-to-end at the application server—will detect tampering, since the OTP-based MIC must match exactly.

VII. Conclusion

This paper underscores the critical role of wearable devices in delivering continuous, reliable healthcare monitoring within energy-scarce environments. By examining LoRaWAN's strengths and constraints—particularly its low-power efficiency and known security shortcomings—a holistic framework has been introduced to bolster data confidentiality and integrity without introducing prohibitive computational or battery overhead. The combined use of EDHOC for key exchange, dynamic session key refresh, and per-transmission tokenization addresses key vulnerabilities including replay, eavesdropping, and bit-flipping attacks.

These findings highlight that LoRaWAN, when secured through lightweight mechanisms, can offer a viable option for large-scale patient monitoring in remote or disaster-stricken regions, where high-power protocols are neither feasible nor affordable. Future work will focus on evaluating battery performance, measuring transmission overhead, and conducting real-world threat simulations to validate the efficacy and scalability of the proposed approach. Building on this foundation can contribute to establishing a new standard for dependable, resilient, and energy-efficient health monitoring solutions.

Acknowledgment

Supervised by Dr. Mekouar Loubna. All authors contributed equally to this work.

References

[1] S. Abdulmalek, A. Nasir, and W. A. Jabbar, "LoRaWAN-based hybrid internet of wearable things system implementation for smart healthcare," *Internet of Things*, vol. 25, 2024, p. 101124.

[2] J. Kang and X. Yeo, "Advances in Energy Harvesting Technologies for Wearable Devices," *(Unpublished/Conference/Journal details)*.

[3] H. Noura, et al., "LoRaWAN security survey: Issues, threats and possible mitigation techniques," *Internet of Things*, vol. 12, 2020.

[4] LoRa Alliance, "LoRaWAN 1.1 Specification," 2017. https://resources.lora-alliance.org/technical-specifications/lorawan-specification-v1-1

[5] R. Sanchez-Iborra, et al., "Enhancing LoRaWAN security through a lightweight and authenticated key management approach," *Sensors*, vol. 18, no. 6, p. 1833, 2018.

[6] A. Kumar, et al., *Wireless Networking*, Elsevier, 2008.

[7] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed., Prentice Hall, 2002.

[8] N. Sornin, et al., "LoRaWAN Specification," LoRa Alliance, 2015.

[9] J. Kim and J. Song, "A secure device-to-device link establishment scheme for LoRaWAN," *IEEE Sens. J.*, vol. 18, no. 5, 2018.

[10] G. Avoine and L. Ferreira, "Rescuing LoRaWAN 1.0," *Financial Cryptography and Data Security*, Springer, 2018.

[11] J. Kim and J. Song, "A simple and efficient replay attack prevention scheme for LoRaWAN," *ICCNS*, 2017.

[12] X. Yang, et al., "Security vulnerabilities in LoRaWAN," *IoTDI*, IEEE, 2018.

[13] G. Selander, et al., “Ephemeral Diffie-Hellman Over COSE (EDHOC),” RFC 9528, IETF, 2023.

[14] L. Lamport, “Password authentication with insecure communication,” *Communications of the ACM*, vol. 24, no. 11, 1981.