International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 24s, 2025
https://theaspd.com/index.php

Social Network Security and Cybersecurity: A State-of-the-Art Survey of Challenges and Emerging Threats

Vishal Vikram Singh[1], Bineet Kumar Gupta[2], Satya Bhushan Verma[3], Veena Singh[4], Rohit Singh[5]
[1,2,4,5]Shri Ramswaroop Memorial University, Barabanki, India 225003
[3]Faculty of Engineering and Technology, University of Lucknow, Lucknow, India, 226007
vishal29.singh@gmail.com, bkguptacs@gmail.com, satyabverma1@gmail.com,
drveenasingh27@gmail.com, rohitsingh.ids@srmu.ac.in

Abstract:
Social networks have transformed the way individuals communicate, share information, and build digital communities. However, this widespread connectivity has introduced significant cybersecurity concerns that threaten user privacy, data integrity, and system resilience. This state-of-the-art survey explores the evolving landscape of social network security, identifying core challenges such as account hijacking, misinformation dissemination, malware propagation, identity theft, and AI-driven threats. The paper critically analyzes existing security frameworks, emerging attack vectors, and current defense mechanisms deployed across various platforms. In addition, it highlights the role of artificial intelligence, blockchain, and regulatory policies in shaping the future of secure social networking. By synthesizing recent advancements and ongoing research, this survey aims to provide a comprehensive understanding of the current threat environment and proposes a strategic outlook toward mitigating emerging risks in social network ecosystems.
Keywords: *Cyber Security, Social Network Security, Challenges*

1. INTRODUCTION

Social networks have evolved into powerful platforms for the rapid dissemination of information, offering users unrestricted access to a vast array of news, updates, and opinions across diverse subjects [1]. Their growing popularity is driven by changing user behavior, technological innovations, and evolving social norms [2]. At the core of these platforms lies a sophisticated network architecture, where social network applications are directly linked to centralized service providers that facilitate communication, data storage, and user interaction [3].

In today's digitally connected world, social applications function as critical nodes within this network, relying on centralized entities that host and manage the underlying infrastructure supporting global social media ecosystems. Users from around the world connect to these centralized servers to interact, share content, and collaborate—generating dynamic exchanges of data that form the foundation of digital communication [4]. This continuous flow includes the transmission of messages, multimedia, and other data, weaving a complex digital fabric of user relationships. The centralized service provider plays an essential role in orchestrating this digital ecosystem, ensuring the stability, responsiveness, and overall performance of social applications as users navigate and engage within this interconnected environment.

1.1. Cyber Threats and their Impact on Social Network Applications

As the number of users and organizations on social networking platforms continues to rise, there is a corresponding increase in the presence of malicious actors targeting these networks. When various social applications connect to a centralized system, this system governs and manages them using specific algorithms to ensure secure service delivery to users. However, despite these mechanisms, social networks remain vulnerable to different types of attacks. On average, approximately 70% of attacks in social network applications are Sybil attacks, 18% are botnet attacks, and 12% are anomaly-based intrusions.

Consider a hypothetical social network with both internal and external architecture, as illustrated in Fig. 1. Within this environment, organizations offer services to users through social applications, where users are connected via routers and switches. Some users are genuine, while others aim to exploit the system and steal data. Internally, two prevalent types of attacks are identified: **Sybil attacks**, where attackers attempt to gain unauthorized access to user profiles through network infrastructure and inject malicious traffic, and **spammer attacks**, which involve distributing harmful messages to multiple users in an effort to extract sensitive information.

Externally, attackers may attempt to compromise the social network through Distributed Denial of Service (DDoS) attacks, the injection of malicious traffic, or other strategies designed to undermine network

security. The core objective is to effectively distinguish between legitimate and malicious users to maintain a secure and trustworthy digital environment.

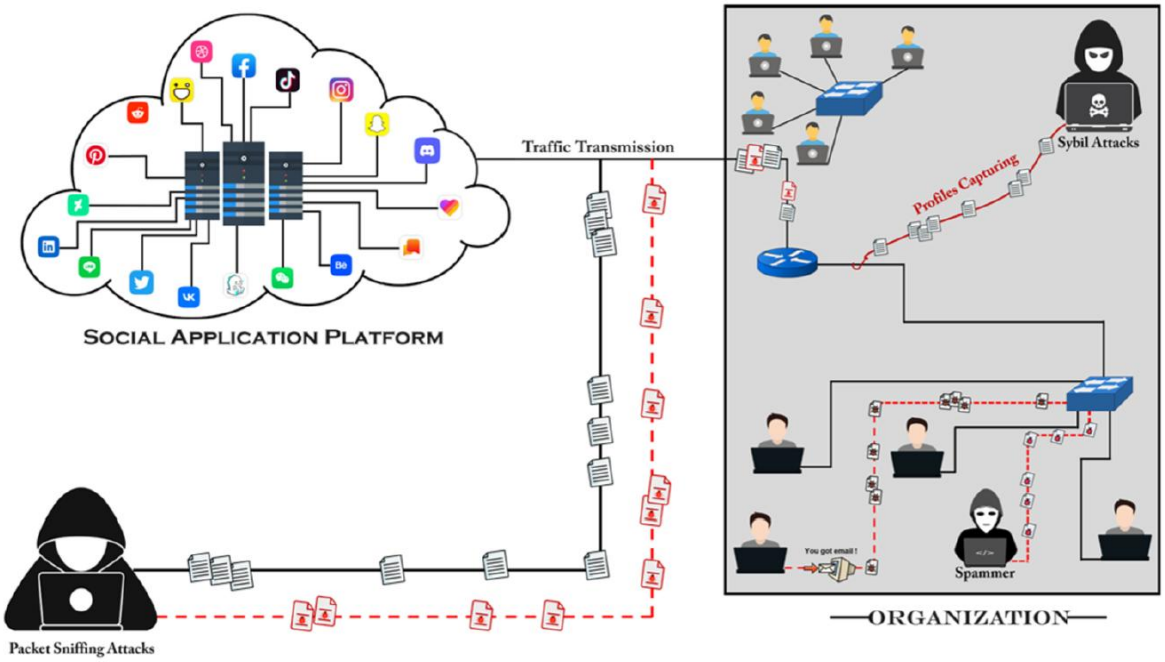

Figure 1: External and Internal attack vectors in Social Networks [50].

The " $T(u)$" represents the traffic generated by user " u ", while "$\sigma(u)$" indicates the probability that, user " u " is harmful. A Sybil attack often generates detrimental traffic across routers. Let " $T(u)$" represent the traffic flow:

$$T(u) = \int f_\tau(u,t)dt \quad (1)$$

Where;

$U = (u_1, u_2, \ldots\ldots, u_n)$: Represent the complete set of users in the social network.

$A = (a_1, a_2, \ldots\ldots, a_m)$ "Represent a collection of defined attack vectors $(a_1 = \text{Sybil}, a_2 = \text{Spammer}, a_3 = \text{DDOS}, a_4 = \text{Malicious Injection})$

"$f_\tau(u,t)$" denotes the traffic flow rate of user " u" at a time " t ". The duration for observation is "τ". If the traffic flow "T(u)" exceeds a threshold "T_{th}" the user is flagged for additional scrutiny.

$$\text{FlagSybil: } T(u) > T_{th} \quad (2)$$

When spammers send irregular message rates across the network, it may be denoted as " $M(u)$":

$$\text{FlagSpammer: } M(u) > M_{th} \quad (3)$$

When we assess the probability of a user. The probability of a user exhibiting malevolence is a weighted sum "$\sigma(u)$" of their actions throughout all attack vectors:

$$\sigma(u) = ([w_1 * p_{sybil}(u)] + [w_2 * p_{spam}(u)] + [w_3 * p_{DDOS}(u)] + [w_4 * p_{MI}(u)] \quad (4)$$

"w_i" signifies the weight of attack type "i" based on its impact.

$p_{sybil}(u)$, $p_{spam}(u)$, $p_{DDOS}(u)$ and $p_{MI}(u)$ denotes the probability of user u being a target.

Consequently, the probability may be calculated as follows:

$$P_{Attack \to (u)} = \frac{\text{Observed behavior for Attack}}{\text{Attack Threshold}}$$

We analyse the fluctuations in user-generated traffic over a defined observation period to identify potential DDoS attacks and malicious traffic injections. The traffic variance, represented as $\text{Var}(T(u))$, is calculated as follows:

$$\text{Var}(T(u)) = \int_0^\tau (f_\tau(u,t) - \mu_\tau(u))^2 \quad (5)$$

Where:

$f_\tau(u,t)$: Traffic flow rate of user u at a time i.

$\mu_\tau(u)$: Mean traffic rate for u during the observation window τ.

$$\text{FlagMalicious} : \text{Var}(T(u) > \text{Var}_{th}$$

This method may detect abnormal and irregular traffic patterns often linked to such attacks.

$$\text{Authenticate } (u \in U) \text{if } (\sigma(u)) < \sigma_{th} \quad (6)$$

Upon integrating all components, the model analyses traffic " $T(u)$", messaging " $M(u)$", and variance $\text{Var}(T(u)$ against thresholds and computes $\sigma(u)$ for user classification, as delineated in equation (6). Nonetheless, the obstacles and concerns encountered by users on the social network.

1.2. Core Threats and Challenges in Social Media Security

Researchers have proposed several algorithms and frameworks aimed at enhancing the security of social networks. However, many of these solutions possess inherent limitations that restrict their effectiveness in the dynamic, complex landscape of social networking platforms.

Challenge 1: Current research lacks a comprehensive framework capable of simultaneously detecting, classifying, and recognizing a wide range of cyberattacks within social networks. Moreover, only about 10–15% of existing studies utilize large-scale public or crowdsourced datasets for training, which limits the generalizability and robustness of their models.

Challenge 2: There is a pressing need for a framework that incorporates multiple community-based models, where each community is composed of agents trained on distinct tasks. However, most existing research does not adequately support this multi-agent, community-driven architecture.

Challenge 3: No existing framework integrates multiple large language models (LLMs), each trained on different datasets, to collaboratively address diverse security threats. Furthermore, current models lack the adaptive capability that would allow agents to assume each other's responsibilities in the event of failure—an essential mechanism for responding dynamically to evolving attacker strategies.

Challenge 4: A truly secure social networking structure—where agents, LLM communities, and algorithms are independently trained but function collaboratively—has yet to be developed. Present frameworks fall short of delivering a cohesive, multilayered prevention system capable of neutralizing threats at the network level during active attacks.

Challenge 5: Most current frameworks lack the flexibility to dynamically adapt to new and evolving threats. They do not support the seamless integration of new agents or communities in response to changes in attack patterns.

Challenge 6: To date, no social network security framework has undergone rigorous validation using statistical, probabilistic, and experimental methods. This gap leaves the actual reliability and real-world effectiveness of existing solutions largely unverified.

These six challenges reflect critical shortcomings in current social network security research and represent key threats that continue to compromise platform integrity. To date, no singular framework has been proposed that holistically addresses all of these issues. However, the proposed **Magteon Turing L3TM** framework aims to overcome these limitations by introducing a unified, adaptive, and resilient solution to fortify social network environments against a broad spectrum of cyber threats.

Let us explain how all these issues might disrupt the functioning of social network security. We possess a social network and have created a unified paradigm for it. Consider we have a model $M(x; \theta)$ in which θ denotes the model parameters. We train model "M" on a single public or crowdsourced dataset "D".

$$D = \{(x_i, y_i)\}_{i=1}^{N}$$

When a model is trained on a dataset, the raw data is organized by the use of Feature Extraction $\phi(x)$, followed by model training to achieve the accuracy "A". When we possess restricted models M and datasets D and we train the model on tiny datasets D_{train} excessive variance may manifest in the loss function. When a model is trained on inadequate data, it often memorizes the training instances rather than learning generalizable patterns, resulting in high accuracy on the training data "$A_{[train]} \rightarrow 1''$ but low accuracy on the test data, leading to overfitting "$A_{[test]} << A_{[train]}$". We need a singular model that has capabilities for User behavior identification "B", Attack detection "D", and Attack classification "X". Furthermore, augmenting "$|D_{train}|$" by data enhancement or synthetic data generation is essential. Utilize " $R(\theta)$" to improve generalization, optimize "MI(f,y)" to ensure substantial features, and use association rules to bolster resilience against unforeseen occurrences.

2. Related Works

Review of Related Work in Social Network Security, The proliferation of abusive activities across social network platforms continues to pose serious security concerns. Instagram, among others, has been a focal point of numerous studies aimed at combating such threats.

Wu et al. [5] proposed a multidimensional feature extraction framework integrated with Support Vector Machines (SVM), achieving an impressive detection accuracy of 99.8%. However, the widespread use of Virtual Private Networks (VPNs) complicates attacker identification due to traffic encryption and obfuscation.

To address encrypted traffic challenges, the Flowlike approach [6] was introduced, attaining 99.2% accuracy in anomaly detection within encrypted communications. Similarly, Bakhshi et al. [7] developed a hybrid deep learning model combining Convolutional Neural Networks (CNNs) and Gated Recurrent Units (GRUs) to distinguish between malicious and benign traffic, significantly improving detection effectiveness.

The complexity introduced by techniques like VPNs and Tor, which modify packet headers, was tackled through AI-FlowDet [8], which extracts 294 statistical and structural (S&S) features, achieving 98.5% accuracy in encrypted environments. Meanwhile, ENiD [9] utilized four machine learning models to identify spoofed encrypted web traffic, yielding an F1 score and accuracy of 97%. This study analyzes and contrasts various container architectures and their configurations within micro-hosting environments [40].

Recognizing the privacy risks of metadata analysis in encrypted environments, Kour et al. [10] used XGBoost to differentiate VPN from non-VPN traffic, achieving 92.4% accuracy across multiple public datasets. Extending beyond traffic analysis, Kour et al. [11] also proposed a Depression Detection Framework using Twitter data, predicting mental health indicators with 94.28% accuracy.

In behavior-based traffic analysis, Wu et al. [12] developed BehaveSniffer, a Graph Convolutional Network (GCN)-based system leveraging Traffic Burst Graphs (TBGs), reaching 99.8% accuracy. To overcome the limitations of basic statistical techniques, Wang et al. [13] used Graph Neural Networks (GNNs) on attack graphs, achieving 99% accuracy on three benchmark datasets [49].

Zhou et al. [14] introduced a 1D-CNN model enhanced with normalization and attention mechanisms, extracting features from hexadecimal data and achieving 98.8% accuracy. To address the limitations of single-modal classifiers, FusionTC [15] employed a two-layer stacking classifier to combine multi-model features.

In encrypted traffic anomaly detection, Long et al. [16] applied L1 regularization for feature selection and achieved a near-perfect 99.98% accuracy. Addressing multimedia threats in smart cities, [17] presented a hybrid model combining GoogLeNet with GNNs, classifying cyberbullying across text, image, and video with 96% accuracy.

Bot detection was tackled by Bazm et al. [18], who used behavioral features and an AdaBoost classifier on a labeled dataset of 2,000 accounts, achieving 95% accuracy. For spam detection, [19] proposed a popularity-based approach utilizing Particle Swarm Optimization (PSO) for feature selection, achieving 99.5% accuracy. CNN-based supervised clustering for bot detection was introduced by Wanda et al. [20], who optimized performance using polling layers.

Privacy-centric frameworks were explored in [21], where a decentralized encryption system using RSA and AES was proposed for key management. Dewan and Kumaraguru [22] designed a two-stage harmful content filter combining URL blacklists and classifiers (RF, SVM, NB), with an average accuracy of 80%. Sen and Aggarwal [23] focused on fake like detection, while Rathore et al. [24] proposed SpamSpotter, evaluated across eight classifiers, with Naïve Bayes yielding 98.4% accuracy. Kiran et al. [25] developed a behavioral risk assessment model for fake account detection on Twitter, validated using real-world data.

Using RapidMiner Studio, [26] applied supervised and unsupervised algorithms to Facebook data, with supervised clustering achieving 97% accuracy. Hakimi et al. [27] used a four-cluster system with KNN, NN, and SVM, where KNN performed best at 82% accuracy.

Sybil detection remains vital. Researchers in [28] created a dataset of 995 images and used AdaBoost, achieving a 99% F1 score. Akyon and Kalfaogh [29] developed dual-labeled datasets, reporting 94% accuracy. Sheikhi [30] compared Bagging classifiers with five traditional models, reaching 98% accuracy. Munoz and Pinto [31] introduced a novel fake profile dataset with a 96% false positive detection rate.

Saranya et al. [32] combined NLP and ML techniques using SVM and Naïve Bayes to detect fake Instagram accounts, achieving 91.5% accuracy. Meshram et al. [33] evaluated Instagram behavior features with Neural Networks, SVM, and Random Forest, obtaining 97% accuracy.

Islam et al. [34] explored deep learning for Malware Intrusion Detection (MID), citing its scalability and efficiency. In [35], user data-sharing behavior was studied with respect to age, gender, and privacy concerns. Agrawal et al. [36] proposed a taxonomy for encrypted mobile traffic classification, while Boutaba [37] analyzed machine learning in networking, discussing its potential and limitations.
Recent research has shifted focus to Large Language Models (LLMs) in social networking environments. Zeng et al. [38] discussed LLM deployment challenges and development concerns. Das et al. [39] analyzed privacy and security risks in LLMs, including vulnerabilities in education, public transit, and healthcare applications.

3. Emerging Threats and Security Challenges in Social Networks

3.1. Identification of External Threats: Behavior Analysis and Traffic Patterns

Researchers have explored a wide range of algorithms and techniques to detect external threats in social networks by analyzing user behavior and traffic patterns. While progress has been made, a comprehensive review of the literature reveals that detecting anomalous user behavior and malicious traffic remains a major challenge in the field of social network security.
Although numerous models and frameworks have been proposed, a significant limitation persists: most of these models are trained and evaluated using only two or three publicly available datasets. The effectiveness and reliability of detection algorithms are highly dependent on the diversity and quality of the training datasets. Unfortunately, a substantial portion of the current research—over 70%—relies predominantly on datasets from the Canadian Institute of Cybersecurity, particularly the ISCX dataset series, which focuses primarily on three categories of encrypted traffic: VPN, non-VPN, and Tor.
This heavy reliance presents a critical constraint, as VPN traffic, in particular, complicates the identification of malicious behavior due to its encrypted and obfuscated nature. In contrast, non-VPN traffic provides clearer signals, making the detection of harmful activity more feasible. Researchers [56–65] have acknowledged these challenges and consistently sought better methods to distinguish between legitimate and malicious behavior across encrypted and unencrypted environments.
Between 2015 and 2024, advancements in traffic analysis and attacker behavior detection evolved significantly. Prior to 2016, security efforts largely centered around website-based applications. However, with the shift in user preferences and service delivery mechanisms, social apps have become the primary medium of interaction. This change has led to a substantial increase in both user engagement and the frequency of cyberattacks post-2016.
To address these evolving threats, researchers introduced numerous algorithms targeting traffic classification and behavioral analysis. Our research synthesizes a decade of developments (2016–2024) and compiles effective strategies into a comparative graphical analysis. This trend highlights the growing interest in understanding user interaction patterns within social networks.
Despite the progress, evaluating the effectiveness of different methodologies remains a challenge due to inconsistencies in experimental setups and visual representations across studies.
In the context of social networks, attacks can arise in two primary forms: external attacks, originating from outside the platform, and internal attacks, executed by compromised or malicious users within the network. Our objective is to determine the authenticity of users under both scenarios. When a user interacts with a social application, advanced algorithms must assess behavior in real time to differentiate between genuine users and potential threats, enabling proactive blocking of malicious actors.
In this section, we propose a novel model aimed at accurately distinguishing legitimate users from inauthentic ones, leveraging traffic analysis and behavior monitoring. This model is a key component of the Magteon Turing L3TM framework, which is designed to overcome the limitations of current approaches by integrating advanced detection techniques with broader, more diverse data inputs.

3.2. Addressing Internal Security Threats: Sybil Attacks, Bots, and Anomalies

Ensuring the security of social networks and applications requires robust defense mechanisms against a variety of internal threats [51]. Among these, bot attacks, Sybil attacks, and anomalies represent the most persistent and damaging intrusions [54-56]. As noted by researchers [52], these threats severely compromise the authenticity and reliability of social platforms, making it critical to distinguish between genuine and malicious users [53].
To combat such threats, researchers have developed numerous techniques aimed at detecting and classifying malicious activities within social applications [57]. The primary objective of these algorithms is to accurately identify inauthentic behavior and reduce the incidence of internal attacks on these

platforms. After proposing various detection methodologies, researchers typically evaluate their effectiveness using statistical and comparative analyses against existing solutions. These studies consistently report improvements in detection accuracy and classification performance.

Despite the existence of many effective algorithms, limitations remain. A truly reliable algorithm for securing social networks must address all categories of internal threats, adapt to diverse attack environments, and sustain performance under real-world conditions. An algorithm's validity is determined not only by its precision but also by its adaptability and resilience.

Although bot, anomaly, and Sybil attacks are classified differently, they are closely related. Attackers frequently create fake identities to deploy bots, a tactic characteristic of Sybil attacks [57]. These bots are then used to carry out malicious activities targeting legitimate users and networks [58].

A common form of anomaly attack involves unexpected or suspicious communications received by users. The decision to accept or reject these interactions can have significant consequences, as such messages often aim to manipulate or deceive the recipient [59]. Addressing these challenges effectively could improve trust in social networks by over 80%, according to several studies [60].

Each year, new algorithms are introduced with claims of outperforming previous models in detecting internal threats. However, the dynamic and evolving nature of cyberattacks—including phishing, spam, fake profile infiltration, and behavioral manipulation—demands more than isolated algorithmic improvements. The mechanisms employed by attackers are constantly shifting, necessitating adaptive and intelligent solutions.

To that end, there is a growing need for a secure, scalable architecture capable of learning from extensive and diverse datasets. The implementation of community-based agents, where each agent is specialized and trained on real-time data, offers a promising path forward. These agents can be collectively optimized to handle various attack types based on their behavioral signatures [61].

This research introduces the Magteon-Turing L3TM framework as a comprehensive solution to these challenges. Unlike previous approaches, no existing algorithm or published method fully addresses the multi-dimensional threats of bots, Sybil attacks, and anomalies in an integrated and adaptive manner. The proposed framework aims to fill this critical gap by leveraging modular, agent-driven intelligence trained on dynamic, real-world datasets [62].

4. Real-World Application Areas of Cybersecurity

Cybersecurity ensures the integrity, confidentiality, and trustworthiness of systems by implementing intrusion detection, defense mechanisms, and encryption technologies. As both social interactions and industrial operations increasingly rely on networked services, cybersecurity has become essential across numerous domains. Two critical application areas include smart grid systems and vehicular communication networks.

4.1. Cybersecurity in Smart Grid Systems

Smart grids represent the next generation of power systems, combining advanced communication and computing technologies to enhance energy efficiency, reliability, and responsiveness. These grids integrate distributed intelligence, demand-side management, and renewable energy sources to deliver improved energy solutions. By reducing latency and enabling faster responses, smart grids provide consumers with more efficient and personalized services [41].

However, due to the extensive interconnectivity of digital devices and communication networks within electrical infrastructure, smart grids are highly vulnerable to cyber threats. The communication networks within smart grids serve as mission-critical platforms for exchanging data across the energy infrastructure [42]. To safeguard these systems, cybersecurity must ensure core objectives such as trust, integrity, and confidentiality.

Achieving these goals requires a multi-layered defense approach that includes:

- Strong authentication mechanisms
- Secure communication protocols
- End-to-end data encryption
- Continuous network monitoring for detecting anomalies

These measures help ensure that only authorized entities can access or manipulate sensitive information, without compromising its security or availability.

- Cybersecurity in smart grids is also vital for mitigating specific threats, such as:
- Denial-of-Service (DoS) attacks
- Advanced Persistent Threats (APTs)
- Unauthorized data breaches

Without adequate security protocols, attacks on smart grids could lead to large-scale power outages, economic disruptions, or even threats to public safety [43]. Strengthening cybersecurity in this sector not only involves deploying technical solutions but also calls for:

- Regular vulnerability assessments
- Enhanced incident response strategies
- Workforce training on cybersecurity best practices
- Greater collaboration between utility providers, regulators, and cybersecurity experts

4.2. Cybersecurity in Vehicular Communication Systems

In modern automotive ecosystems, cybersecurity is fundamental to enabling secure vehicular communication and protecting critical in-vehicle and network infrastructure. Vehicles increasingly interact with each other (V2V), with infrastructure (V2I), and with broader networks (V2X), contributing to improved road safety, optimized traffic flow, and enhanced passenger experience [44].

This communication network involves several components, such as:

- Electronic Control Units (ECUs)
- Sensors and in-vehicle software
- Roadside units and control centers

Cybersecurity in this context must safeguard both digital and physical vehicle components, as well as sensitive passenger data—including location, driving behavior, and health-related information.

Maintaining secure and reliable vehicle communication systems is especially critical in emergency scenarios, where the timely delivery of safety messages can prevent accidents or save lives. Therefore, routine system diagnostics, timely software updates, and redundancy mechanisms are essential components of a robust cybersecurity strategy in the automotive domain.

Furthermore, protecting the vehicular communication ecosystem requires:

- Secure firmware updates and patch management
- Intrusion detection systems for vehicle networks
- Strong data privacy safeguards
- Resilience against unauthorized access and manipulation

To ensure long-term cybersecurity in this field, a collaborative and forward-thinking approach is necessary. Key stakeholders—including automotive manufacturers, network providers, government agencies, service operators, and end-users—must work together to establish standards, share threat intelligence, and continuously adapt to evolving risks [45].

4.3. Cybersecurity in Smart Cities

A smart city is an urban environment that leverages advanced technologies and communication networks to enhance the quality of life for its citizens. However, the increasing dependence on interconnected devices and systems brings with it substantial cybersecurity risks. Addressing these risks is essential to safeguard citizens' privacy, ensure public safety, and maintain the resilience of critical infrastructures [46].

Cybersecurity in smart cities focuses on protecting key infrastructure—such as energy grids, water supply systems, transportation networks, and communication services—as well as securing Internet of Things (IoT) devices like sensors, surveillance cameras, and actuators. It also ensures the protection of sensitive personal information and disaster recovery systems, which are vital in times of crisis.

4.4. Cybersecurity in Smart eHealth Systems

IoT-enabled healthcare solutions—such as remote patient monitoring and smart health applications—rely extensively on internet-connected devices to gather health-related data from sources like wearable devices, medical instruments, and mobile applications. The integration of IoT with medical systems enhances

healthcare service quality by providing real-time monitoring, early intervention, and continuous progress tracking.

However, the healthcare sector has become a frequent target of cyberattacks. According to the World Economic Forum, over 10 million records have been compromised globally, including patient health data, HIV test results, and provider information. In some incidents, attackers accessed more than 3 million records across 33 countries [47].

Cybersecurity in this domain is essential to protect sensitive medical data, ensure secure communication between patients and providers, and maintain the integrity and reliability of healthcare services. Implementing strong encryption, authentication protocols, and continuous monitoring can mitigate these risks effectively.

5. Challenges in Cybersecurity

In the digital era, cybersecurity poses major challenges for individuals, corporations, and governments. As the use of connected devices and data-driven systems grows, so does the complexity of protecting them from threats. Despite technological advancements, cybercriminals are evolving their techniques, making it increasingly difficult to detect and defend against attacks [48].

5.1. Evolving Nature of Cyberattacks

Modern cyber threats are sophisticated and often difficult to detect. Attackers now employ advanced strategies such as:

Advanced Persistent Threats (APTs): Long-term, targeted attacks often supported by governments or organized cybercriminals. They focus on sectors like defense, IT, and national security. Example: Operation Aurora (2009), which targeted companies like Google and Adobe.

Zero-Day Exploits: These attacks exploit unknown software vulnerabilities before developers have a chance to issue a fix. Example: Stuxnet, which used several zero-day flaws to disrupt Iran's nuclear program.

5.2. IoT Security Challenges

The growing number of Internet of Things (IoT) devices introduces unique vulnerabilities:

Limited Security: IoT devices often have constrained computing resources, limiting their ability to support robust security mechanisms.

Outdated Firmware: Many IoT products are shipped with outdated or weak security configurations.

Privacy Concerns: Devices collect sensitive data—such as location, biometrics, and behavior patterns—which must be properly encrypted and securely stored to prevent unauthorized access.

5.3. AI-Driven Cyberattacks

Artificial Intelligence (AI) and Machine Learning (ML) are now being exploited by attackers:

AI-Assisted Attacks: AI is used to automate and optimize attack strategies.

AI-Driven Autonomous Attacks: Fully automated cyberattacks carried out by AI systems.

Deepfakes: AI-generated fake content used for misinformation, fraud, or impersonation.

AI-Enhanced Botnets: More agile and stealthy botnets capable of large-scale DDoS attacks and data theft.

Reinforcement Learning Attacks: These adapt over time to evade traditional security measures.

5.4. Cloud Security Issues

As organizations increasingly rely on cloud services, new vulnerabilities arise:

Data Breaches: Exposing confidential data stored in the cloud.

Weak Access Control: Insecure credentials can lead to unauthorized access.

Insecure APIs: Exploitable interfaces used to access cloud services.

Service Downtime: Interruptions affecting availability and business operations.

6. Common Cyber Threats and Attacks

Eavesdropping: Intercepting private communication (emails, calls, traffic) without permission—also known as sniffing or snooping.

Traffic Analysis: Monitoring communication patterns to gather metadata and launch further attacks.

Replay Attacks: Capturing and re-transmitting valid data to trick systems into unauthorized operations.

Man-in-the-Middle (MiTM): Interception and manipulation of communications between two parties, often over insecure networks.
Impersonation: Masquerading as trusted individuals or organizations (e.g., phishing or social engineering) to steal sensitive data.
Denial-of-Service (DoS/DDoS): Overwhelming systems with traffic to make services unavailable. DDoS uses multiple devices to amplify the attack.
Malware: Malicious software like viruses, worms, ransomware, and spyware used to disrupt or steal data.
Scripting Attacks: Including SQL injection and XSS, used to manipulate or steal data from vulnerable web applications.
Insider Threats: Employees or users with elevated access misuse their privileges to harm the organization or leak sensitive data.
Physical Theft of Smart Devices: Stolen IoT or smart devices can leak private data via techniques like power analysis.
Birthday Attack: Exploits weaknesses in hashing algorithms to find two different inputs with the same hash, compromising password systems.
Dictionary Attack: Attempts to crack passwords by trying commonly used or simple combinations from a list.
Stolen Verifier Attack: Stealing password verifiers stored on systems to impersonate users.
Session Key Computation Attack: Targeting cryptographic session keys used between users to decrypt or modify communication.

6.1. Attacks on Machine Learning (ML) Models

Machine learning systems are also vulnerable to several types of attacks:
Dataset Poisoning: Introducing malicious samples into training data to alter ML behavior.
Model Poisoning: Manipulating ML model parameters to reduce accuracy or induce errors.
Privacy Breach: Leaking of training data or model internals due to poor encryption or storage.
Runtime Disruption: Targeting deployed ML systems through attacks like phishing, SQL injection, or DoS to halt their operation.
Mitigation strategies include the use of blockchain for decentralized training, differential privacy techniques (e.g., noise injection), and secure, encrypted environments for ML deployment.

CONCLUSION:

In the rapidly evolving digital ecosystem, social networks have become indispensable platforms for communication, collaboration, and commerce. However, their expansive growth has simultaneously introduced a myriad of cybersecurity challenges and emerging threats. This survey underscores the pressing need to address vulnerabilities such as data breaches, identity theft, misinformation campaigns, and AI-driven social engineering attacks. The integration of advanced technologies like blockchain, machine learning, and privacy-preserving models presents promising avenues for fortifying these platforms. Nevertheless, the dynamic nature of threats calls for continuous innovation, multi-stakeholder collaboration, and robust policy enforcement. Strengthening social network security is no longer a technical choice but a strategic imperative to preserve user trust, data integrity, and the overall stability of the digital society.

REFERENCES:

[1]. Y. Liu, J. Huang, Y. Li, D. Wang, B. Xiao, Generative AI model privacy: a survey, Artif. Intell. Rev. 58 (1) (2024) 33, https://doi.org/10.1007/s10462-024-11024-6.
[2]. Q. Lai, H. Hua, Secure medical image encryption scheme for Healthcare IoT using novel hyperchaotic map and DNA cubes, Expert. Syst. Appl. 264 (2025) 125854,https://doi.org/10.1016/j.eswa.2024.125854.
[3]. R. Chapaneri, S. Shah, Enhanced detection of imbalanced malicious network traffic with regularized generative adversarial networks, J. Netw. Comput. Appl. 202(2022) 103368, https://doi.org/10.1016/j.jnca.2022.103368.
[4]. A. Najafi, O. Varol, TurkishBERTweet: fast and reliable large language model for social media analysis, Expert. Syst. Appl. 255 (2024) 124737, https://doi.org/10.1016/j.eswa.2024.124737.
[5]. H. Wu, Q. Wu, G. Cheng, S. Guo, Instagram user behavior identification based on multidimensional features, in: IEEE INFOCOM 2020 - IEEE Conference onComputer Communications Workshops (INFOCOM WKSHPS), IEEE, 2020, pp. 1111–1116, https://doi.org/10.1109/INFOCOMWKSHPS50562.2020.9162688.
[6]. T. Shapira, Y. Shavitt, FlowPic: a generic representation for encrypted traffic classification and applications identification, IEEE Trans. Netw. Serv. Manag. 18(2) (2021) 1218–1232, https://doi.org/10.1109/TNSM.2021.3071441.
[7]. T. Bakhshi, B. Ghita, Anomaly detection in encrypted internet traffic using hybrid deep learning, Secur. Commun. Netw. 2021 (2021) 1–16, https://doi.org/10.1155/2021/5363750.

[8]. H.-Y. Chen, T.-N. Lin, The challenge of only one flow problem for traffic classification in identity obfuscation environments, IEEE Access. 9 (2021)84110–84121, https://doi.org/10.1109/ACCESS.2021.3087528.
[9]. G. Mengmeng, Y. Xiangzhan, V.Mysore Sachidananda, L. Shangqing, L. Likun, ENiD: an encrypted web pages traffic identification based on web visiting behavior, in: 2022 IEEE International Conference on Data Mining Workshops (ICDMW), IEEE, 2022, pp. 593–601, https://doi.org/10.1109/ICDMW58026.2022.00082.
[10]. Z. Wang, B. Ma, Y. Zeng, X. Lin, K. Shi, Z. Wang, Differential preserving in XGBoost model for encrypted traffic classification, in: 2022 International Conference on Networking and Network Applications (NaNA), IEEE, 2022, pp. 220–225, https://doi.org/10.1109/NaNA56854.2022.00044.
[11]. H. Kour, M.K. Gupta, An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM, Multimed. Tools. Appl. 81 (17) (2022) 23649–23685, https://doi.org/10.1007/s11042-022-12648-y.
[12]. T. Wu, et al., BehavSniffer: sniff user behaviors from the encrypted traffic by traffic burst graphs, in: 2023 20th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), IEEE, 2023, pp. 456–464, https://doi.org/10.1109/SECON58729.2023.10287511.
[13]. L. Wang, Z. Cheng, Q. Lv, Y. Wang, S. Zhang, W. Huang, ACG: attack classification on encrypted network traffic using graph Convolution attention Networks, in: 2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD), IEEE, 2023, pp. 47–52, https://doi.org/10.1109/CSCWD57460.2023.10152599.
[14]. Y. Zhou, et al., Identification of encrypted and malicious network traffic based on one-dimensional convolutional neural network, J. Cloud Comput. 12 (1) (2023) 53, https://doi.org/10.1186/s13677-023-00430-w.
[15]. S. Li, et al., FusionTC: encrypted app traffic classification using decision-level multimodal fusion learning of flow sequence, Wirel. Commun. Mob. Comput. 2023 (2023) 1–15, https://doi.org/10.1155/2023/9118153.
[16]. G. Long, Z. Zhang, Deep encrypted traffic detection: an anomaly detection framework for encryption traffic based on parallel automatic feature extraction, Comput. Intell. Neurosci. 2023 (1) (2023), https://doi.org/10.1155/2023/3316642
[17]. S. R, A. J, Classification of cyberbullying messages using text, image and audio in social networks: a deep learning approach, Multimed. Tools. Appl. 83 (1) (2024) 2237–2266, https://doi.org/10.1007/s11042-023-15538-z.
[18]. M. Bazm and M. Asadpour, Behavioral modeling of Persian Instagram users to detect bots. 2020. doi:10.48550/arXiv.2008.03951.
[19]. [22] S.R. Sahoo and B.B. Gupta, "Popularity-based detection of malicious content in Facebook using machine learning approach," 2020, pp. 163–176. doi:10.1007/978-981-15-0029-9_13.
[20]. P. Wanda, M.E. Hiswati, H.J. Jie, DeepOSN: bringing deep learning as malicious detection scheme in online social network, IAES Int. J. Artif. Intell. 9 (1) (2020) 146, https://doi.org/10.11591/ijai.v9.i1.pp146-154.
[21]. H. Jashn, B. Mahipour, E. Moharamkhani, B. Zadmehr, A framework for privacy and security on Social networks using encryption algorithms, Int. J. Smart Electr. Eng. (2023) 31–41.
[22]. P. Dewan, P. Kumaraguru, Facebook Inspector (FbI): towards automatic real-time detection of malicious content on Facebook, Soc. Netw. Anal. Min. 7 (1) (2017) 15, https://doi.org/10.1007/s13278-017-0434-5.
[23]. I. Sen, A. Aggarwal, S. Mian, S. Singh, P. Kumaraguru, A. Datta, Worth its weight in likes, in: Proceedings of the 10th ACM Conference on Web Science, ACM, New York, NY, USA, 2018, pp. 205–209, https://doi.org/10.1145/3201064.3201105.
[24]. S. Rathore, V. Loia, J.H. Park, SpamSpotter: an efficient spammer detection framework based on intelligent decision support system on Facebook, Appl. Soft. Comput. 67 (2018) 920–932, https://doi.org/10.1016/j.asoc.2017.09.032.
[25]. K. Kiran, C. Manjunatha, T.S. Harini, P.Deepa Shenoy, K.R. Venugopal, Identification of anomalous users in twitter based on user behaviour using artificial neural networks, in: 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), IEEE, 2019, pp. 1–5, https://doi.org/10.1109/ I2CT45611.2019.9033728.
[26]. M. Albayati, A. Altamimi, MDFP: a machine learning model for detecting fake facebook profiles using supervised and unsupervised mining techniques, Int. J. Simul. (2020), https://doi.org/10.5013/IJSSST.a.20.01.11.
[27]. A.N. Hakimi, et al., Identifying Fake Account in Facebook Using Machine Learning, 2019, pp. 441–450, https://doi.org/10.1007/978-3-030-34032-2_39.
[28]. Y. Singh, S. Banerjee, Fake (Sybil) account detection using machine learning, SSRN Electronic J. (2019), https://doi.org/10.2139/ssrn.3462933.
[29]. F.C. Akyon, M.Esat Kalfaoglu, Instagram fake and automated account detection, in: 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), IEEE, 2019, pp. 1–7, https://doi.org/10.1109/ASYU48272.2019.8946437.
[30]. S. Sheikhi, An efficient method for detection of fake accounts on the Instagram platform, Rev. d'Intell. Artif. 34 (4) (2020) 429–436, https://doi.org/10.18280/ria.340407.
[31]. S.D. Munoz, E. Paul Guillen Pinto, A dataset for the detection of fake profiles on social networking services, in: 2020 International Conference on Computational Science and Computational Intelligence (CSCI), IEEE, 2020, pp. 230–237, https://doi.org/10.1109/CSCI51800.2020.00046.
[32]. S. Saranya Shree, C. Subhiksha, R. Subhashini, Prediction of fake instagram profilesusing machine learning, SSRN Electron. J. (2021), https://doi.org/10.2139/ssrn.3802584.
[33]. Er.P. Meshram, R. Bhambulkar, P. Pokale, K. Kharbikar, A. Awachat, Automaticdetection of fake profile using machine learning on Instagram, Int. J. Sci. Res. Sci.Technol. (2021) 117–127, https://doi.org/10.32628/IJSRST218330.
[34]. M.R. Islam, S. Liu, X. Wang, G. Xu, Deep learning for misinformation detection on online social networks: a survey and new perspectives, Soc. Netw. Anal. Min. 10 (1) (2020) 82, https://doi.org/10.1007/s13278-020-00696-x.
[35]. Mashael Aljohani, Alastair Nisbet, Kelly Blincoe, A survey of social media users privacy settings & information disclosure, in: Australian Information Security Management Conference, 2016, pp. 67–75.
[36]. A. Agrawal, et al., A survey on analyzing encrypted network traffic of mobile devices, Int. J. Inf. Secur. 21 (4) (2022) 873–915, https://doi.org/10.1007/s10207-022-00581-y.
[37]. R. Boutaba, et al., A comprehensive survey on machine learning for networking: evolution, applications and research opportunities, J. Internet Serv. Appl. 9 (1) (2018) 16, https://doi.org/10.1186/s13174-018-0087-2.

[38]. Jingying Zeng, Richard Huang, Waleed Malik, Langxuan Yin, Bojan Babic, Large language models for social networks: applications, challenges, and solutions, ArXiv. (2024).
[39]. Badhan Chandra Das, M.Hadi Amini, Yanzhao Wu, Security and privacy challenges of large language models: a survey, ArXiv. (2024).
[40]. SB Verma, Brijesh P., and BK Gupta, Containerization and its Architectures: A Study, ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, Vol. 11 N. 4 (2022), 395-409, eISSN: 2255-2863, DOI: https://doi.org/10.14201/adcaij.28351
[41]. Anamika Agarwal, S. B. V., B. K. Gupta, A Review of Cloud Security Issues and Challenges, ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, Issue, Vol. 12 N. 1 (2023), pp 1-22, eISSN: 2255-2863, 2023 https://doi.org/10.14201/adcaij.31459
[42]. A. Abraham et al., "Naïve Bayes Approach for Word Sense Disambiguation System With a Focus on Parts-of-Speech Ambiguity Resolution," in IEEE Access, vol. 12, pp. 126668-126678, 2024, doi: 10.1109/ACCESS.2024.3453912
[43]. S. Singh, S. B. V, Resolving Covid-19 with Blockchain and AI: A Systematic Review, ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, Vol. 13 (2024), pp 1-18,eISSN: 2255-2863, 2024 https://doi.org/10.14201/adcaij.31454
[44]. Satya B Verma, Shashi B V, Data Transmission in BPEL (Business Process Execution Language), ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal Regular Issue, Vol. 9 N. 3 (2020), 105-117 eISSN: 2255-2863 DOI: https://doi.org/10.14201/ADCAIJ202093105117 105
[45]. B.A. Scott, M.N. Johnstone, P. Szewczyk, S. Richardson, BGP anomaly detection as a group dynamics problem, Comput. Netw. 257 (2025) 110926, https://doi.org/10.1016/j.comnet.2024.110926.
[46]. C. Jisi, B. Roh, J. Ali, An effective scheme for classifying imbalanced traffic in SDIoT,leveraging XGBoost and active learning, Comput. Netw. 257 (2025) 110939, https://doi.org/10.1016/j.comnet.2024.110939.
[47]. X. Chen, et al., High-performance routing for hose-based VPNs in multi-domain backbone networks, Comput. Netw. 57 (4) (2013) 944–953, https://doi.org/10.1016/j.comnet.2012.11.010.
[48]. C. Xenakis, C. Ntantogian, I. Stavrakakis, A network-assisted mobile VPN for securing users data in UMTS, Comput. Commun. 31 (14) (2008) 3315-327, https://doi.org/10.1016/j.comcom.2008.05.018.
[49]. S. Lv, C. Wang, Z. Wang, S. Wang, B. Wang, Y. Zhang, AAE-DSVDD: a one-class classification model for VPN traffic identification, Comput. Netw. 236 (2023) 109990, https://doi.org/10.1016/j.comnet.2023.109990.
[50]. Muhammad Nadeem, Advancing social network security with magteon-turing L3TM: A multi-layered defense system against cyber threats, Computer Networks, 267 (2025), https://doi.org/10.1016/j.comnet.2025.111375
[51]. J. Li, B. Feng, H. Zheng, A survey on VPN: taxonomy, roles, trends and future directions, Comput. Netw. 257 (2025) 110964, https://doi.org/10.1016/j.comnet.2024.110964.
[52]. K. Raghavan, et al., Advancing anomaly detection in computational workflows with active learning, Fut. Gener. Comput. Syst. 166 (2025) 107608, https://doi.org/10.1016/j.future.2024.107608.
[53]. B. Bertalani˘c, V. Han˘zel, C. Fortuna, Explainable semantic wireless anomaly characterization for digital twins, Comput. Netw. 251 (2024) 110660, https://doi.org/10.1016/j.comnet.2024.110660.
[54]. N. Nafti, O. Besbes, A. Ben Abdallah, A. Vacavant, M.H. Bedoui, A fast residualattention network for fine-grained unsupervised anomaly detection and localization, Appl. Soft. Comput. 165 (2024) 112066, https://doi.org/10.1016/j.asoc.2024.112066.
[55]. S. Corli, L. Moro, D. Dragoni, M. Dispenza, E. Prati, Quantum machine learning algorithms for anomaly detection: a review, Fut. Gener. Comput. Syst. 166 (2025) 107632, https://doi.org/10.1016/j.future.2024.107632.
[56]. K. Liu, et al., SFACIF: a safety function attack and anomaly industrial condition identified framework, Comput. Netw. 257 (2025) 110927, https://doi.org/10.1016/j.comnet.2024.110927.
[57]. A.J. Hashim, M.A. Balafar, J. Tanha, A. Baradarani, Adaptive deep learning models for efficient multivariate anomaly detection in IoT infrastructures, Appl. Soft. Comput. 167 (2024) 112377, https://doi.org/10.1016/j.asoc.2024.112377.
[58]. F. Marulli, P. Paganini, F. Lancellotti, The three sides of the moon LLMs in cybersecurity: guardians, enablers and targets, Procedia Comput. Sci. 246 (2024) 5340–5348, https://doi.org/10.1016/j.procs.2024.09.653.
[59]. S. Shafee, A. Bessani, P.M. Ferreira, Evaluation of LLM-based chatbots for OSINTbased Cyber threat awareness, Expert. Syst. Appl. 261 (2025) 125509, https://doi.org/10.1016/j.eswa.2024.125509.
[60]. J. Liu, G. Lin, H. Mei, F. Yang, Y. Tai, Enhancing vulnerability detection efficiency:an exploration of light-weight LLMs with hybrid code features, J. Inf. Secur. Appl.88 (2025) 103925, https://doi.org/10.1016/j.jisa.2024.103925.
[61]. K. Suresh, K. Jayasakthi Velmurugan, R. Vidhya, S. Rahini sudha, Kavitha, Deep anomaly detection: a linear one-class SVM approach for high-dimensional and large-scale data, Appl. Soft. Comput. 167 (2024) 112369, https://doi.org/10.1016/j.asoc.2024.112369.
[62]. A. Iliopoulos, J. Violos, C. Diou, I. Varlamis, Feature bagging with nested rotations (FBNR) for anomaly detection in multivariate time series, Fut. Gener. Comput. Syst. 163 (2025) 107545, https://doi.org/10.1016/j.future.2024.107545.