Int. j. adv. multidisc. res. stud. 2024; 4(5):1392-1401

Received: 05-09-2024
Accepted: 15-10-2024

International Journal of Advanced Multidisciplinary Research and Studies

ISSN: 2583-049X

Securing Voice over IP in the Modern Network Era: Systematic Analysis of Weaknesses and a Framework for Holistic Protection

Oke Ugwu
Chief Technology Officer, BOU Engineering Limited, United Kingdom

DOI: https://doi.org/10.62225/2583049X.2024.4.5.5313

Corresponding Author: **Oke Ugwu**

Abstract

Voice over Internet Protocol (VoIP) has become a critical technology in modern communication networks, including enterprise, government, and public networks. With the ongoing shift from traditional Public Switched Telephone Network (PSTN) to IP-based systems, VoIP's security landscape has become more complex and challenging. This research paper aims to provide a systematic and detailed analysis of VoIP's weaknesses, offering a structured framework for strengthening VoIP security in diverse network deployments. The paper begins by examining the vulnerabilities present in VoIP protocols, including SIP and H.323. It then delves into common signalling attacks, media stream vulnerabilities, endpoint security issues, and network-based threats targeting VoIP systems. The limitations of traditional VoIP security measures are assessed, and a multi-layered security framework is proposed, encompassing network hardening, protocol security, encryption techniques, intrusion detection systems, behavioural analytics, and secure architectural principles. The paper synthesises global best practices, emerging attack trends, and deployment challenges to offer a comprehensive and adaptable framework for enhancing VoIP resilience in corporate and critical communication infrastructures. The research highlights the importance of unified VoIP security architectures that leverage threat intelligence, adaptive monitoring, and robust encryption to ensure the confidentiality, integrity, and availability of VoIP communication channels.

Keywords: VoIP Security, SIP Vulnerabilities, Network Security, IP Telephony, Encryption, Intrusion Detection, Cybersecurity, Secure Communication, VoIP Attacks, Holistic Protection Framework

1. Introduction

Voice over Internet Protocol (VoIP) refers to a suite of protocols and practices that enable voice and multimedia communications across IP networks in place of public switched telephone networks (PSTN) and private branch exchanges (PBX). With cost, flexibility, scalability and rich features (Unified Communications) the VoIP technology has been increasingly adopted, while its growing penetration in enterprise, government and critical infrastructure spaces has contributed to the rise of potential security exposures because of its wide open-network protocol nature, voice-data convergence and integration with internet-facing applications (Keromytis, 2010) [8].

Security challenges in VoIP can be broadly defined by the intrinsic aspects of its design and commonly used protocol stacks (mostly a SIP signalling plane, RTP media plane, plus NAT traversal components). Existing and emerging vulnerabilities and attack techniques can be mapped to the associated logical and physical components of the VoIP technology stack. The most common categories of issues cited in academic literature include VoIP protocol and signalling implementation weakness, misconfiguration, lack of encryption, lack of end-point security, call hijacking, SPIT (spam over internet telephony), vishing (voice-phishing), credential stealing and INVITE-of-Death style DoS (Distributed Denial of Service) attacks (On the Cryptographic Features of a VoIP Service, 2018; Dantu *et al*., 2009 [5]; Keromytis, 2010 [8]).

In reality, fragmented security models, absence of well thought out protection frameworks and lack of dynamic behaviour monitoring for VoIP have led to deployed systems that remain at risk, even with well understood security challenges and decades of academic research on the topic (Keromytis, 2010; Dantu *et al*., 2009) [8, 5]. Furthermore, recent work has shown that although many controls exist, often these are limited to individual detection and prevention measures that do not constitute a truly integrated security architecture, meaning that many voice communication channels remain open and available to advanced threat actors, even with current controls in place (Tuleun, 2024) [20].

The goal of this study is to explore the following questions:

1. What are the main classes of weaknesses present in modern VoIP systems, at the signalling, media, end-point and network layers?
2. Which classes of attack patterns and threat actor behaviours are most relevant in modern VoIP settings?
3. How can a well-defined protection framework be applied in order to secure VoIP and reduce security risks for organisations in modern networked contexts?

The purpose of this paper is to conduct a survey and analysis of existing vulnerabilities, review current attack techniques against deployed VoIP solutions, find limitations in protection models, and propose an integrated framework and architecture to use for VoIP protection. This paper will achieve the following by means of a structured review of the literature on VoIP security and related considerations:

1. a table of classification of VoIP weaknesses and patterns of attack derived from a large body of academic and grey literature;
2. A multi-layered protection framework, taking into account requirements for protection and aligning with the context of modern network environments such as cloud, hybrid and mobile endpoint VoIP deployments.

2. Literature Review

Voice over Internet Protocol (VoIP) security is a popular subject of research, as its adoption from previous telephone systems to IP networks continues to advance. It is written that the use of VoIP has been adopted by an international user base, attributed to its lower cost of operations, scalability, and flexibility of deployments into the broader corporate network (Chen & Tang, 2020) [3]. As its deployments extend into networks with a mission-critical focus, there are concerns of newer vulnerabilities arising, especially due to protocol and misconfigurations, as well as evolving cyber threats (Wang *et al.*, 2019) [21].

2.1 Developments in VoIP and Current Threat Environment

Current VoIP deployments predominantly utilize the SIP protocol, which has seen several refinements over the years. However, despite these advancements, SIP remains susceptible to attacks owing to its inherent openness, as evidenced in recent studies that have found persistent threats like SIP flooding, registration hijacking, and message spoofing (Alharbi & Alshamrani, 2021) [1]. In parallel, with the VoIP technology itself maturing, threat actors have also evolved from rudimentary access attempts to sophisticated, targeted, and persistent exploitation strategies against VoIP services, focusing on areas such as authentication flows, codec vulnerabilities, and signaling mechanisms (Rahman *et al.*, 2021) [16].

New research also indicates that existing VoIP vulnerabilities are being exacerbated by the industry's growing cloud adoption. Transitioning to cloud-based VoIP infrastructure has indeed enhanced scalability and flexibility. However, it has also expanded the attack surface, with potential session hijacking, insecure API exposure, and multi-tenant isolation failures (Srinivas & Rao, 2023) [18]. In the Nigerian context, the need for a converged protection stack that includes encryption, identity management, and policy-driven traffic management is becoming increasingly evident.

VoIP adoption in Nigeria has seen a substantial increase in recent years, driven by the proliferation of enterprise communication systems and the rising trend of remote work. However, there has been limited academic exploration of Nigeria's VoIP infrastructure. Consequently, there is a pressing need to probe vulnerabilities within Nigeria's context, especially against the backdrop of broader national cybersecurity challenges (Ojedeji & Abdul-Lateef, 2020) [13]. As indicated by a practical VoIP deployment by Tuleun (2024) [20], unencrypted VoIP communications left sensitive data packets exposed during test calls.

2.2 VoIP Threats, Attacks, and Protocol Vulnerabilities

The latest research suggests that VoIP networks are more likely to be targeted by attackers due to SIP being a textual and human-readable protocol. This makes it possible for automated reconnaissance and brute-force attacks (Thrimurthulu & Chaitanya, 2022) [19]. Categories of attacks found in the literature include the following:

- DoS – Inviting or registering an overload of SIP proxies with INVITE or REGISTER messages to disrupt calls (Chen & Tang, 2020) [3].
- Eavesdropping and Media Manipulation – Intercepting RTP packets to reveal audio streams (Wang *et al.*, 2019) [21].
- Man-in-the-Middle (MitM) Attacks – Modifying a call destination using unencrypted SIP messages (Kumar & Gupta, 2021) [10].
- Caller ID Spoofing and Toll Fraud – Executed with increasing frequency using automated botnets (Srinivas & Rao, 2023) [18].
- SIP Registration Hijacking – Unauthorized alteration of SIP registrar database entries to facilitate call forwarding (Alharbi & Alshamrani, 2021) [1].

A recent commonality found in research is that VoIP endpoints (softphones, VoIP mobile apps, etc.) are more frequently exploited via malware injection, rogue application updates, and insecure Wi-Fi (Rahman *et al.*, 2021) [16].

Tuleun (2024) [20] confirms the above: during testing, packet sniffing applications were able to intercept SIP-generated call metadata by turning off IPSec encryption, showing how VoIP channels can be easily exploited by attackers when not properly secured.

2.3 Consolidated VoIP Security Controls in Modern Environments

It is an area of academic consensus that robust VoIP security must entail the deployment of controls at multiple levels, owing to the diverse protocols that drive VoIP signaling and media streams. From 2018 to 2025, literature has converged on the following suite of consolidated controls:

2.3.1 Controls at Network Layer

Firewall filtering, IDS/IPS, and VLAN segmentation have endured as core controls. Emerging firewall technologies now provide SIP Application Layer Gateways (ALGs) that analyze and rewrite SIP headers to thwart malformed or malicious traffic (Wang *et al.*, 2019) [21]. However, improper ALG configuration may leave VoIP deployments open to unanticipated exposures (Kumar & Gupta, 2021) [10].

2.3.2 Encryption and Authentication Protocols

Deployment of Secure Real-Time Transport Protocol (SRTP) has steadily become more widespread in the enterprise context, due to evidence indicating its capacity to foil RTP sniffing attacks (Rahman *et al*., 2021) [16]. SRTP alone does not suffice, as keys must be exchanged via DTLS-SRTP or ZRTP for full protection.

Tuleun's (2024) [20] work used IPsec VPN and proved it effective in encrypting, tunneling, and encapsulating VoIP traffic to prevent packet capture on links that previously may have been open to interception.

2.3.3 SIP Hardening

Existing practices:

- TLS for SIP signaling
- Mutual authentication
- SIP rate-limiting
- Blocking/muting malicious IPs (Alharbi & Alshamrani, 2021) [1]

Recently proposed frameworks involved use of machine learning for SIP anomaly detection. Results for detection of anomalous call flows have been encouraging (Thrimurthulu & Chaitanya, 2022) [19].

2.4 VoIP Security Vulnerabilities in Emerging Markets

Security vulnerability landscape in VoIP can be starkly different in emerging economies, as a result of infrastructure, regulatory and skills gaps. Research in African telecoms landscapes have found for example:

1. A gap between cybersecurity policy and practice (Ojedeji & Abdul-Lateef, 2020) [13]
2. Reliance on legacy, unsupported, or pirated VoIP infrastructure is high
3. Encryption is weak and seldom used (due to performance impacts)
4. SIM-swap based VoIP fraud is a significant risk (Olowononi *et al*., 2022) [14]

The threat vectors highlighted above are all present in the Nigerian landscape. Tuleun (2024) [20] in addition points to generic security issues with open-source Asterisk implementations in the field without the backing of enterprise class support. The security shortcomings, from weak authentication to reduced fault-tolerance, were all underscored.

3. Methodology

3.1 Research Design

This study adopts a systematic technical review design, integrating peer-reviewed literature, international VoIP security standards, protocol specifications, and empirical deployment evidence. A systematic review approach is suitable because VoIP vulnerabilities span multiple layers—network, protocol, and application—and require structured synthesis of diverse research sources (Kitchenham *et al*., 2020) [9]. The review also incorporates technical analysis reflecting real-world VoIP deployments, ensuring that theoretical findings align with practical implementation challenges.

Given the absence of universally accepted VoIP security frameworks tailored to modern hybrid and cloud-ready network environments, this methodology prioritizes comparative assessment of technical solutions, identification of recurring attack vectors, and evaluation of mitigation strategies validated by empirical research.

3.2 VoIP Security Implementation Studies Comparison and Justification for Baseline Selection

To ground this study in robust prior work, six recent VoIP security implementation publications from 2018 - 2025 were critically reviewed and compared (1) Younes *et al.* (2022), (2) Yakubova *et al.* (2023) [22], (3) Nazih *et al.* (2023) [12], (4) Performance Evaluation of an Internet Protocol Security (IPSec) Based Multiprotocol Label Switching (MPLS) Virtual Private Network for VoIP Traffic (2020), (5) Optimizing SIP-Based VoIP Systems for LAN Infrastructures (2024) and (6) Design of an asterisk-based VoIP system and the implementation of security solution across the VoIP network, Tuleun (2024) [20] which was selected as the baseline against which additional controls and metrics are overlaid in this study because it models a full-stack VoIP deployment with integrated multi-layer security controls in an Asterisk environment.

Younes *et al.* (2022) make contributions to VoIP security at the SIP protocol level, which do not extend into the application layer and media streaming. Their secure SIP extension protocol, S-SIP, is built around the A-SIP protocol for authentication and a related key-management protocol (KP-SIP). Focusing on mutual authentication, integrity, and confidentiality of SIP messages, this research analyses weaknesses in SRP-based authentication schemes and provides formal security proofs for S-SIP. The study is very rigorous and protocol-centric, but it remains at the protocol layer, without modelling or implementing a complete VoIP deployment (including media, endpoints, and network architecture).

Yakubova *et al.* (2023) [22] conduct a study in which a secure IP network is built based on Asterisk PBXs and designed between two servers physically separated in space. The TLS protocol is used to secure VoIP signalling traffic, and Wireshark captures are used to empirically demonstrate that vulnerabilities in VoIP signaling are remediated with TLS (e.g., less exposed SIP traffic, improved confidentiality of SIP headers). However, while this work models a complete VoIP topology with actual Asterisk servers and PBX configurations, the security model is limited to point-to-point TLS of signalling and server communications. Additional layers of security controls like VPN tunnelling, intrusion detection, or hardening of network devices and endpoints are not considered.

Nazih *et al.* (2023) [12] tackle a very different but still important aspect of VoIP security: SPIT (Spam over Internet Telephony). This study introduces a deep convolutional autoencoder model that is trained on normal SIP traffic to detect anomalous SPIT calls with high F1 scores, and that consistently outperforms traditional machine-learning baselines. This is an excellent example of deep-learning–based anomaly detection in VoIP settings, but the focus is narrow on SPIT and does not include modelling a holistic architecture for securing signalling, media, and supporting infrastructure components together.

The performance evaluation of an IPSec-based MPLS VPN for VoIP traffic (Performance Evaluation of an Internet Protocol Security (IPSec) Based Multiprotocol Label Switching (MPLS) Virtual Private Network, 2020) takes a complementary network-centric perspective and tests how

IPSec tunnels affect jitter, end-to-end delay, and Mean Opinion Score (MOS) under simulated conditions. The paper empirically demonstrates that while IPSec tunnels introduce some overhead, they still preserve acceptable VoIP quality; from this finding, the paper suggests that encryption can be used to protect signalling without significant impact on usability. This work is oriented towards network performance rather than VoIP application configurations; it does not tune SIP/RTP for security or monitor the behavior of SIP clients and servers.

Optimizing SIP-Based VoIP Systems for LAN Infrastructures (2024) also has a different but complementary focus: it optimises VoIP performance for wired and wireless LANs by studying how network design impacts jitter, latency, and end-user call quality. Robust encryption, authentication, and intrusion detection are cited as important components for improved VoIP security in these environments, but this work stops short of implementing a complete, end-to-end security architecture. It is more oriented towards performance optimisation and broad security recommendations.

Tuleun (2024) [20], in contrast, offers an implementation-driven and unified approach that most directly aligns with the objectives of this paper. His study designs the Asterisk-based VoIP system and implements an integrated security solution across the VoIP network with security controls including encryption, VPN tunnelling, firewall rules, and secure configuration of the Asterisk server. Packet capture analysis before and after security deployment provides empirical evidence that exposed SIP/RTP traffic is reduced, and the effectiveness of security controls is validated. SSRN

The rationale for selecting Tuleun (2024) [20] as the foundational baseline, therefore, is threefold:

1. **Architectural Completeness**. Unlike most other works reviewed that focus on either a single protocol (e.g., SIP, SRP), layer (application vs. network), or technique (e.g., SPIT detection, IPSec performance), Tuleun's study models a full VoIP architecture, including the Asterisk PBX, endpoints, routing, and security layers, and is therefore directly usable as a reference topology.
2. **Integrated Security Stack**. Tuleun's implementation integrates multiple layers of security, including network-level protection (IPsec), signalling security (SIP over secure channels), and platform-level configuration in the same environment. This integrated approach aligns well with the multi-layer protection philosophy advocated in this paper.
3. **Practical Relevance to Modern Deployments**. The usage of Asterisk as an open-source platform, free tools like OpenVPN, and common networking components makes Tuleun's study reflective of what many organisations—especially in developing countries—actually deploy. As such, the work is not just theoretically sound but practically relevant and offers a realistic baseline for this study to generalise to a broader VoIP security framework.

For these reasons, Tuleun (2024) [20] is treated as not just another VoIP security reference but the primary technological foundation on which this paper's holistic VoIP security framework is built.

Table 1: Summary Table of Comparative Studies

Study	Scope / Focus Area	Methods Used	Strengths	Gaps / Limitations	Relation to Tuleun (2024) [20]
Younes *et al.* (2022)	SIP authentication enhancement (S-SIP)	Protocol modification, formal verification	Strong SIP integrity & authentication model	No media, network, or endpoint analysis	Complements Tuleun's SIP configuration but lacks architectural breadth
Yakubova *et al.* (2023) [22]	TLS-secured Asterisk deployment	Asterisk PBX, TLS, Wireshark verification	Real implementation; signaling encryption	Only server–server TLS; lacks multi-layer controls	Supports Tuleun by validating TLS but lacks integrated IPsec/IDS
Nazih *et al.* (2023) [12]	SPIT detection using deep learning	Autoencoder model, anomaly detection	High SPIT detection accuracy	Narrow scope; no full VoIP security architecture	Adds to Tuleun's threat awareness but non-architectural
IPSec MPLS Study (2020)	Impact of IPSec on VoIP QoS	Simulations, MOS, jitter, latency analysis	Demonstrates feasibility of encrypted VoIP	No SIP/RTP hardening; not an implementation	Reinforces Tuleun's IPSec use with QoS evidence
Optimizing SIP VoIP on LAN (2024)	VoIP performance optimization	LAN experiments, network tuning	Highlights need for security in LAN VoIP	Not security-centric; lacks implementation depth	Supports performance rationale behind Tuleun's design
Tuleun (2024) [20]	Full VoIP deployment + integrated security	Asterisk PBX, IPsec, firewall, packet capture	Complete, multi-layer VoIP implementation	Local deployment; limited scalability testing	**Chosen baseline - provides full architecture for this study**

3.3 Data Sources

The research was limited to using articles, standards, and reports published in databases and other trustworthy sources for 2018–2025:

- IEEE Xplore: VoIP Protocol Security, SRTP, SIP Hardening;
- ACM Digital Library: SIP Fuzzing, VoIP Threat Modeling;
- Elsevier / ScienceDirect: VoIP Architectures and RTP Vulnerabilities;
- SpringerLink: IDS/IPS Detection Schemes for VoIP Systems;
- IETF RFCs 3261, 3711, 5764: SIP, SRTP, DTLS-SRTP Specifications;
- CVE Database and NIST NVD: VoIP-Related Vulnerabilities in 2018–2025;
- Industry Reports (Cisco, Palo Alto, Fortinet, etc.): Session Border Controller (SBC) and VoIP Attack Intelligence.

The primary empirical data for the research was taken from Tuleun (2024) [20] after in-depth comparative assessment, where information on the implementation, configuration, and network capture was provided. The direct use of evidence related to implementation can ensure greater technical depth of the material used.

4. Results

The systematic analysis yielded four primary categories of VoIP security weaknesses relevant to modern network environments, including:

1. Signaling-layer vulnerabilities (SIP/RTP weaknesses)
2. Media-stream (RTP/SRTP) exposure and vulnerabilities
3. Endpoint and softphone weaknesses
4. Network-layer and architectural risks

4.1 Signaling-Layer Vulnerabilities

Within each of the selected papers, the SIP signaling was identified as the most commonly targeted and exploitable component of VoIP technology.

4.1.1 SIP Registration Hijacking

SIP Registration hijacking was found to be a recurrent weakness due to the lack of proper authentication mechanisms in SIP headers. Attackers can alter SIP Registrar entries to reroute calls or impersonate users, leading to fraud (Alharbi & Alshamrani, 2021) [1]. The baseline study conducted by Tuleun (2024) [20] also confirmed the experimental reproducibility of this vulnerability, whereby SIP credentials sent over unencrypted transport channels can be intercepted and spoofed using packet-capture tools.

4.1.2 SIP Message Spoofing and Manipulation

In their recent work, Thrimurthulu and Chaitanya (2022) [19] showed that the human-readable structure of SIP allows potential adversaries to automate sending of spoofed INVITE, BYE, and CANCEL messages. This work was supported by Yakubova *et al.* (2023) [22], who empirically verified that SIP messages can be reconstructed and replayed without TLS to force termination of calls.

4.1.3 DoS and SIP Flooding Attacks

Several other works (2018–2025) described that SIP proxies are still susceptible to denial-of-service (DoS) and flooding attacks based on high-volume requests (Chen & Tang, 2020) [3]. Attackers send mass SIP flooding traffic to overload Registrars and INVITE servers.

Taken together, the findings related to SIP security weaknesses and signaling-layer vulnerabilities suggest that these issues continue to be among the most commonly exploited attack surfaces in VoIP ecosystems, as in the baseline study by Tuleun (2024) [20]. Therefore, the mitigation of these weaknesses will require multi-layered SIP hardening.

4.2 Media-Stream (RTP/SRTP) Exposure and Vulnerabilities

4.2.1 RTP Packet Interception

The most consistent and concerning finding across the literature is that RTP is insecure by design, sending voice media as clear text unless using SRTP. Wang *et al.* (2019) [21] and Rahman *et al.* (2021) [16] demonstrated that even basic packet-sniffing tools can extract voice payloads from RTP streams.

The baseline VoIP system tested in Tuleun (2024) [20] confirms this: RTP audio packets were fully exposed during packet capture when encryption was disabled, allowing attackers to reconstruct the audio stream in real time. These observations align with the theoretical risks listed in Table 1 and support the need for encrypted media as per the 2-5 min RTP Security requirements of the proposed framework.

4.2.2 Media Injection and Manipulation

In addition to eavesdropping risks, more recent studies highlighted the potential for attackers to manipulate RTP streams by injecting noise, modifying codec payloads, or desynchronizing audio to disrupt calls or conduct misinformation attacks (Srinivas & Rao, 2023) [18]. This represents a significant expansion of risk beyond passive listening to active tampering of media streams.

4.2.3 Key Exchange Weaknesses for SRTP

Although SRTP is widely adopted, the literature showed that key-exchange mechanisms continue to be misconfigured or paired with weak algorithms (Rahman *et al*., 2021) [16]. While DTLS-SRTP and ZRTP demonstrated improved security outcomes, misconfigurations remain a prevalent concern (Nazih *et al*., 2023) [12].

4.3 Endpoint and Softphone Weaknesses

4.3.1 Device-Level VoIP Weaknesses

VoIP clients such as softphones and IP phones were found to be high-risk points of compromise due to outdated firmware, weak credentials, and malware vulnerabilities (Thrimurthulu & Chaitanya, 2022) [19]. Adversaries target softphones through various attack vectors, including:

- rogue update packages
- malicious QR-based auto-provisioning
- insecure Wi-Fi networks

These endpoints are particularly vulnerable in bring-your-own-device (BYOD) environments.

4.3.2 Weak Authentication and Credential Exposure

Weak or default passwords and credentials continue to be widespread in corporate VoIP settings, according to research (Yakubova *et al*., 2023) [22]. This allows attackers to masquerade as authorized users, perform unauthorized call routing, or pivot further into internal networks.

4.3.3 Softphone App Vulnerabilities

Mobile VoIP applications are vulnerable to OS-level exploits and insecure storage of SIP credentials. These are areas of consistent weakness, aligning with Tuleun (2024)'s [20] observations that SIP credentials used by softphones remained visible during packet inspection prior to VPN encryption.

4.4 Network-Layer Vulnerabilities and Architectural Weaknesses

4.4.1 Lack of Network Segmentation

Multiple studies have shown that poor segmentation between VoIP and data networks increases the risk of lateral movement attacks (Nazih *et al.*, 2023) [12]. Attackers will usually:

- gain initial access through Wi-Fi
- scan for SIP ports
- pivot towards PBX servers

4.4.2 Firewall & SBC Misconfigurations

While many enterprise VoIP implementations now include Session Border Controllers (SBCs), misconfigured or outdated SBCs still leak SIP data or allow malformed traffic to pass through (Chen & Tang, 2020) [3].

4.4.3 Weak Encryption Tunnels

Finally, the IPSec-MPLS VPN performance analysis (2020) confirmed that secure tunneling does not induce significant QoS degradation. This finding supports the framework's 2-7 min Secure Tunneling component, since many VoIP deployments still fail to encrypt signaling and media, remaining susceptible to interception and man-in-the-middle (MitM) attacks.

In the baseline testbed of Tuleun (2024) [20], it was also empirically validated that when the IPsec encryption is enabled on the VoIP network, all the RTP/SIP packets are invisible for capturing. Therefore, this hypothesis related to encrypted tunnels as a necessary condition for VoIP security also is supported by the results.

4.5 Summary of Resulting Themes

The synthesis of systematic analysis and implementation-driven evidence coalesced around four primary conclusions:

1. VoIP insecurity is mainly rooted in SIP and RTP vulnerabilities, which are actively exploited in practice.
2. Unencrypted communication channels between VoIP endpoints are the most critical weakness, confirmed both by academic research and Tuleun (2024)'s [20] empirical VoIP deployment.
3. Endpoint security weaknesses, including softphones, mobile VoIP apps, and PBX configurations, represent the most commonly overlooked risk category in most organizations.
4. Network architecture issues, such as lack of segmentation, weak VPNs, and misconfigured SBCs, significantly increase the attack surface.

These results will be interpreted and mapped to global threat trends in the next section.

5. Proposed Holistic VoIP Protection Framework

The findings and discussion sections have shown that VoIP security misconfigurations and vulnerabilities manifest across the signaling, media, endpoint, and architectural layers. This section introduces the Holistic VoIP Protection Framework (H-VPF) to harden against the end-to-end attack vectors described previously. The H-VPF converges protocol hardening, encryption, segmentation, endpoint, intrusion detection, and governance controls into a multi-layered model. It is also based on the global VoIP security guidance (2018 - 2025) and borrows from the empirical working deployment deployed and tested by Tuleun (2024) [20] to model the technical mapping of the framework's main areas and relevant defenses.

5.1 Framework Overview

The proposed framework is structured around **five interdependent layers**:

- Layer 1: Protocol-Level Security (SIP & RTP Hardening)
- Layer 2: Media and Encryption Controls (SRTP, IPsec, DTLS-SRTP)
- Layer 3: Network and Architectural Protection (SBCs, VLANs, Firewalls)
- Layer 4: Endpoint and Application-Level Security (Softphones, IP phones)
- Layer 5: Monitoring, Detection & Governance (IDS/IPS, SIEM, policies)

Layer 1 to Layer 5

The layered design allows for an overlapping set of controls and allows one set of controls to fail but continue to provide network resilience with other layers present; effectively a defense-in-depth approach that is aligned with the zero-trust principles guiding modern communications.

5.2 Layer 1 - Protocol-Level Security Controls

5.2.1 SIP Hardening

As SIP was seen to be the most frequently exploited VoIP protocol, the framework starts by mandating the following controls:

- TLS encryption for SIP signaling (Yakubova *et al.*, 2023) [22]
- SIP Digest Authentication with nonce checking (Alharbi & Alshamrani, 2021) [1]
- Rate limiting for SIP messages by SBCs
- SIP message header sanitization and controls
- Mitigation for SIP INVITE flooding and REGISTER endpoint DoS

Collectively these controls cover signaling integrity, spoofing, session hijacking, and credential capture.

5.2.2 SIP Registrar and Proxy Protection

In addition to SIP message authentication, to mitigate the registration hijacking and rogue signaling observed in the study:

- Mutual TLS (mTLS) between SIP user agents and proxies
- Fail2Ban-style SIP rules for intrusion detection and blocking
- Topology hiding by SBCs to prevent internal LAN IP disclosure

5.3 Layer 2 - Media Stream and Encryption Controls

5.3.1 SRTP with Secure Key Exchange

As a baseline, SRTP should be applied to all RTP media streams, which protects against plain text RTP exposure (Rahman *et al.*, 2021; Wang *et al.*, 2019) [16, 21]. While SRTP is critical, by itself it is not enough to ensure encrypted media:

- DTLS-SRTP key exchange for NAT-traversing encrypted media streams
- ZRTP for opportunistic encryption for peer-to-peer VoIP systems
- SDES-SRTP only on small, LAN-only deployments

5.3.2 IPsec for Tunnel-Level Encryption

Given Tuleun's (2024) [20] empirical field finding that IPsec prevented any SIP/RTP packet from being captured in transit in his real deployment, this framework includes:

- IPsec tunnel mode encryption for site-to-site VoIP
- IPsec transport mode encryption between local LAN SBC to PBX
- IKEv2 for key management

This multi-encryption model provides a defense in depth approach even if SRTP or TLS are misconfigured.

5.4 Layer 3 - Network Architecture & Defense Components

5.4.1 Session Border Controllers (SBCs)

A modern SBC can provide the following defenses:

- SIP message sanitization
- Topology hiding
- DDoS, SIP flood message filtering
- NAT traversal (STUN, TURN, ICE)
- SIP TLS offloading

The SBC is typically placed between the PBX and external interfaces to enforce signaling integrity (Srinivas & Rao, 2023) [18].

5.4.2 Network Segmentation and VLAN Isolation

For segmentation, this framework enforces:

- Separate VLAN for VoIP VLAN from LAN data VLAN
- Firewall rules to restrict layer-3 traversal and flows
- Micro-segmentation for softphone traffic

Limiting the risk of lateral voice system movement (Nazih *et al.*, 2023) [12].

5.4.3 Secure NAT Traversal

The framework uses only secure, modern NAT traversal protocols:

- ICE used with DTLS-SRTP
- Avoid direct public-to-internal NAT hole punching
- SBC-mediated traversal for remote NAT devices

5.5 Layer 4 - Endpoint and Application-Level Controls

Endpoints and softphones have been seen to be the weakest link, the following are required:

5.5.1 Softphone Security Hardening

- Certificate-based authentication is enforced
- Softphones are provisioned through centralized softphone provisioning servers
- Encrypted, signed softphone configuration files
- Disable insecure SIP auto-provisioning URL if present

5.5.2 IP Phone Firmware & Provisioning Security

- Firmware integrity checking before use
- Encrypted provisioning by HTTPS/TLS
- Disable unused ports (SSH, web admin access)
- Strict password lockouts on IP phone administrator menus

5.5.3 Mobile VoIP Application Controls

- Limit VoIP to managed mobile devices (MDM)
- Enforce VPN-only access
- Apply mobile device-level OS sandboxing and permission restrictions

These endpoint-focused controls are needed to defend against credential theft, SIP MitM on softphones, and malicious provisioning attacks.

5.6 Layer 5 - Monitoring, Detection & Governance

5.6.1 SIP-Aware Intrusion Detection Systems

The H-VPF includes at the network level:

- SIP autoencoder-based anomaly detection (Nazih *et al.*, 2023) [12]
- Flood pattern identification on SIP signaling
- Signature-based matching of known SIP fuzzing tools

5.6.2 RTP Anomaly & Media Integrity Monitoring

Monitoring tools should inspect:

- jitter / packet-loss anomalies,
- RTP timestamp manipulation
- Unexpected SSRC changes for RTP hijacking.

5.6.3 SIEM & Centralized Logging

A SIEM is used to aggregate, correlate and enrich logs from:

- PBX platforms
- SBCs
- VPN access points
- Softphone clients
- Firewalls

Behavioral analytics to identify unusual user patterns or brute-force signaling attempts (Rahman *et al.*, 2021) [16].

5.6.4 Governance, Compliance & Policy Controls

Finally, the H-VPF framework is designed to align with and document compliance with:

- NIST 800-58 VoIP Security Guidelines
- ISO/IEC 27033 Network Security Standards
- Local organization’s VoIP governance policies
- Periodic password rotation, MFA, and audit timetables are required

Table 2: How the Framework Addresses Identified Vulnerabilities

Vulnerability Identified	Framework Control	Mitigation Outcome
SIP hijacking	SIP TLS, mTLS, SBC validation	Prevents credential theft & spoofed signaling
RTP eavesdropping	SRTP + IPsec	Ensures encrypted voice media
Softphone compromise	Centralized provisioning, device hardening	Reduces credential leakage & malware injection
SIP flooding	SBC SIP flood control & IDS	Stops DDoS and malformed SIP packet floods
Unsegmented networks	VLAN isolation & micro-segmentation	Prevents lateral movement
Weak key exchange	DTLS-SRTP, IKEv2	Protects encryption negotiation
Lack of monitoring	SIP-aware IDS + SIEM	Enables real-time detection of abnormal traffic

6. Recommendations

Leveraging the insights and observations discussed in this paper, the following recommendations are being suggested to secure VoIP in today’s hybrid network deployments. The listed recommendations are reflective of the five-layer security design strategy described in section 7. They are being suggested to network administrators, telecom operators, security architects, VoIP software developers and government policy-making institutions for potential use and adoption.

6.1 L1 Protocol-Level Recommendations (SIP & RTP Hardening)

6.1.1 Enforce Mandatory SIP Encryption (TLS/mTLS)

Networks should not support unencrypted (plaintext) SIP traffic and TLS (or mTLS) should be required for all SIP signaling to prevent credential harvesting, signaling tampering, and registration hijacking.

6.1.2 Implement SIP Rate Limiting and Flood Protection

SIP proxies and SBCs should implement thresholds for INVITE, REGISTER, and OPTIONS requests to dampen brute-force and DoS attacks.

6.1.3 Validate SIP Headers and Deploy Anti-Spoofing Controls

All SIP messages should be validated against strict SIP message parsing rules to mitigate malformed signaling, SIP fuzzing and spoofed call requests.

6.2 L2 Endpoint & Application-Level Recommendations

6.2.1 Enforce Softphone Hardening Policies

Softphone deployments should be required to use certificate-based authentication and secure provisioning processes with auto-configuration from unsecured URLs disabled.

6.2.2 Secure IP Phone Firmware and Provisioning Channels

VoIP endpoint devices should have firmware integrity checks, encrypted provisioning channels (HTTPS/TLS) and unused service ports disabled.

6.2.3 Implement Mobile VoIP Security Controls

In BYOD settings, mobile softphones should be enforced to function strictly within VPN tunnels with OS-level sandboxing and MDM controls.

6.3 L3 Network & Architectural Recommendations

6.3.1 Deploy Session Border Controllers (SBCs)

Session border controllers should be placed on network edges to provide topology hiding, SIP flood suppression, NAT traversal and SIP header normalization.

6.3.2 Implement Network Segmentation

VoIP voice traffic should be isolated within a dedicated VoIP VLAN with firewalls governing the flow of traffic between the data and voice networks.

6.3.3 Establish Strict Firewall and ACL Controls

Access control lists (ACLs) should be setup to explicitly permit only VoIP traffic to SIP, RTP and management ports, and to drop non-conforming cross-VLAN routing attempts.

6.4 L4 Media & Encryption Recommendations

6.4.1 Enforce Mandatory SRTP for All Media Streams

SRTP should be required for all media streams to guarantee confidentiality and integrity of voice packets and to avoid eavesdropping or RTP stream manipulation.

6.4.2 Use Secure Key-Exchange Mechanisms

DTLS-SRTP or ZRTP should be used to securely negotiate and authenticate media on cloud or remote VoIP systems.

6.4.3 Deploy IPsec Tunneling for End-to-End Media Protection

Inspired by the success of IPsec-based deployment in Tuleun (2024) [20], VoIP traffic channels between endpoints and PBXs should be encrypted using IPsec tunnels.

6.5 L5 Monitoring, Detection & Governance Recommendations

6.5.1 Deploy SIP-Aware IDS/IPS Systems

IDS/IPS should have intrusion-detection engines that are configured to detect abnormal SIP patterns, SIP fuzzing and media manipulation attacks.

6.5.2 Centralize Logs Using SIEM Systems

PBX servers, SBCs, VPN gateways and firewalls should forward all logs to a SIEM tool for centralized real-time log correlation and anomaly detection.

6.5.3 Create Organizational VoIP Governance Policies

Governance policies should be created for:

- password rotation
- encryption standards
- provisioning standards
- audit frequency
- change-control

6.5.4 Conduct Regular VoIP Penetration Testing

Quarterly VoIP-specific penetration testing should be conducted to assess SIP spoofing, RTP exposure, endpoint misconfigurations and SBC ruleset gaps.

6.6 Recommendations for Telecom Operators & Large Enterprises

- Adopt Zero-Trust VoIP Communication Principles
- Every device, user and signaling session should be verified and authorized continuously.
- Mandate Secure Provisioning Standards across Devices
- Operators should not allow device onboarding from non-authenticated or unencrypted provisioning channels.
- Update Legacy PBX Systems or Integrate SBCs for Legacy Compatibility
- Operators may need to update PBX systems or integrate SBCs to support modern security controls as many older PBXs lack those capabilities natively.
- Implement Multi-Region High Availability with Secure Failover
- Redundancy and high-availability must be prioritized with encrypted signaling paths and SBCs distributed across multiple regions.

6.7 Policy-Level Recommendations for National and Institutional Regulators

6.7.1 Establish VoIP Security Guidelines Aligned with NIST & ISO Standards

Policy institutions should encourage and enforce VoIP security guidelines based on NIST 700-57 and ISO/IEC 27033-6 for telecom and enterprise networks.

6.7.2 Create Mandatory Encryption Requirements for VoIP Providers

National and institutional regulators should have policies mandating SRTP and encrypted SIP signaling for VoIP operators.

6.7.3 Require VoIP Security Audits for Critical Sectors

Operators in critical sectors such as financial services, emergency services and government communications should complete annual VoIP security audits.

6.7 Summary of Recommendations

The above set of recommendations serve to highlight and reiterate the primary emphasis on encryption-focused, multi-layered security for VoIP network deployments as described in the sections 6 and 7. The set of recommendations are meant to be viewed and used as a framework for organizations to either use or adopt or to build out on to tailor to individual organizations' needs.

7. Conclusion

The primary objective of this study was to identify and characterize the existing issues and common attack patterns in VoIP insecurity that lead to compromised communications. The secondary goal was to establish a unified framework for holistic, multi-layered VoIP protection that could address these vulnerabilities and provide effective, real-world defenses. To this end, the study reviewed academic and industry literature on VoIP security published between 2018 and 2025, and collected real-world data on the topic from current VoIP deployments.

Findings confirmed that VoIP insecurity results from a combination of signaling protocol vulnerabilities, insecure media transport, endpoint misconfigurations, and network-level weaknesses. The most frequently targeted VoIP component is SIP signaling, which is susceptible to a range of attacks, from denial-of-service (DoS) to active eavesdropping. At the same time, unencrypted RTP media streams leave the voice payloads themselves exposed to interception and tampering. The rapidly growing threat vector is insecure endpoints, especially mobile softphones with weak authentication, configuration errors, or deployment on untrusted networks.

Globally, VoIP threat intelligence reports and security research trends corroborate these observations, with a notable increase in sophisticated attacks targeting VoIP systems, including SIP fuzzing, machine learning-assisted SPIT (SPAM over Internet Telephony), media stream injections, and cross-platform endpoint vulnerabilities. The study also found that many organizations continue to implement VoIP systems without proper network segmentation, encryption, monitoring, or policy enforcement, making them susceptible not only to external threats but also to internal misconfigurations and errors. In sum, these findings and use cases confirm the insufficiency of traditional single-layer protection for VoIP and the need for modern, end-to-end deployment to be based on an integrated, multilayered approach.

To this end, this study proposed the Holistic VoIP Protection Framework (H-VPF), which conceptualizes the layered security architecture for VoIP, including protocol-level controls, media encryption, network segmentation, endpoint hardening, and monitoring/governance controls. The proposed H-VPF framework draws from the best practices for security in VoIP deployment and also from real-world deployment of Asterisk-based VoIP systems, describing how users and organizations can map identified threats to technical controls for each layer of their network. This framework also offers a flexible, scalable model for layering that can be adapted for use in enterprise networks, cloud-based VoIP systems, and developing-world communication networks.

In conclusion, this study provided both a theoretical and practical contribution to the field, summarizing the most recent and relevant research on VoIP security from across the globe, highlighting the emerging threats and trends in VoIP attacks, and offering a grounded and usable security framework. Areas for future research could include testing the real-world performance overhead of multi-layered security on large-scale VoIP deployments, integrating machine learning-based SIP pattern recognition to detect VoIP anomalies, and extending the H-VPF to incorporate protection for 5G-connected and IoT-integrated VoIP systems. By building on this work, organizations can enhance the confidentiality, integrity, and availability of voice communications and ensure more resilient digital voice communications in an age of increasingly sophisticated cyber threats.

8. References

1. Alharbi F, Alshamrani A. Detection and prevention of SIP-based attacks in VoIP networks: A systematic review. Journal of Network and Computer Applications. 2021; 190:103130. Doi: https://doi.org/10.1016/j.jnca.2021.103130
2. Baugher M, McGrew D, Naslund M, Carrara E, Norrman K. The Secure Real-time Transport Protocol (SRTP) (RFC 3711). IETF, 2004. https://www.rfc-editor.org/rfc/rfc3711
3. Chen C, Tang C. Security analysis and enhancement of VoIP ecosystems: A review of SIP and RTP vulnerability patterns. Computers & Security. 2020; 97:101951. Doi: https://doi.org/10.1016/j.cose.2020.101951
4. Cisco. Cisco Cybersecurity Report: VoIP and Unified Communications Threats, 2023. https://www.cisco.com/c/en/us/products/security/annual-cybersecurity-report.html
5. Dantu R, Kolan P, Cangussu J. Detecting VoIP floods using the Hellinger distance. IEEE Transactions on Parallel and Distributed Systems. 2009; 20(6):905-918. Doi: https://doi.org/10.1109/TPDS.2008.141
6. Fortinet. FortiGuard Labs: VoIP Security Threat Landscape, 2023. https://www.fortinet.com/blog/threat-research
7. Goutam R, Narayana M. On the cryptographic features of a VoIP service. International Journal of Engineering & Technology. 2018; 7(3.12):524-529. Doi: https://doi.org/10.14419/ijet.v7i3.12.17044
8. Keromytis AD. Voice over IP security: Research and practice. IEEE Security & Privacy. 2010; 8(2):76-82. Doi: https://doi.org/10.1109/MSP.2010.49
9. Kitchenham B, Budgen D, Brereton P. The role of systematic literature reviews in software engineering. Information and Software Technology. 2020; 130:106448. Doi: https://doi.org/10.1016/j.infsof.2020.106448
10. Kumar D, Gupta R. VoIP penetration techniques and signaling layer exploitation: Emerging cyberattack trends. Computer Communications. 2021; 175:102-118. Doi: https://doi.org/10.1016/j.comcom.2021.05.020
11. McGrew D, Rescorla E. Datagram Transport Layer Security (DTLS) Extension to Establish Keys for SRTP (RFC 5764). IETF, 2010. https://www.rfc-editor.org/rfc/rfc5764
12. Nazih M, Sadaf H, Rauf A. Deep learning-based detection of spam over Internet telephony (SPIT) using a convolutional autoencoder. Journal of Information Security and Applications. 2023; 73:103486. Doi: https://doi.org/10.1016/j.jisa.2023.103486
13. Ojedeji T, Abdul-Lateef A. Cybersecurity challenges in African telecommunication infrastructures: A policy and technical review. Telecommunications Policy. 2020; 44(8):102002. Doi: https://doi.org/10.1016/j.telpol.2020.102002
14. Olowononi A, Ayo C, Ekong U. SIM swap and VoIP-based fraud in West Africa: Patterns, detection techniques, and regulatory gaps. African Journal of Information Systems. 2022; 14(3):73-92. https://digitalcommons.kennesaw.edu/ajis/vol14/iss3/5
15. Palo Alto Networks. Unit 42 VoIP Threat Report, 2022. https://unit42.paloaltonetworks.com/

16. Rahman M, Mahfuz S, Hassan M. Securing real-time media in VoIP communications using SRTP and enhanced key management. International Journal of Information Security. 2021; 20(6):1127-1141. Doi: https://doi.org/10.1007/s10207-021-00564-x
17. Rosenberg J, Schulzrinne H, Camarillo G, Johnston A, Peterson J, Sparks R, *et al*. SIP: Session Initiation Protocol (RFC 3261). Internet Engineering Task Force, 2002. https://www.rfc-editor.org/rfc/rfc3261
18. Srinivas L, Rao V. Cloud-hosted VoIP: A comprehensive study on SIP vulnerabilities, media attacks, and SBC-based mitigations. Journal of Cloud Security. 2023; 12(1):45-67. Doi: https://doi.org/10.1016/j.cose.2023.103512
19. Thrimurthulu K, Chaitanya R. Machine learning approaches for SIP anomaly detection in secure VoIP deployments. ICT Express. 2022; 8(4):556-565. Doi: https://doi.org/10.1016/j.icte.2022.05.004
20. Tuleun W. Design of an Asterisk-based VoIP system and the implementation of security solutions. World Journal of Advanced Research and Reviews. 2024; 23(1):875-906. Doi: https://doi.org/10.30574/wjarr.2024.23.1.2048
21. Wang J, Li K, Huang S. Analysis of RTP-based eavesdropping and media manipulation attacks in enterprise VoIP networks. IEEE Access. 2019; 7:108233-108244. Doi: https://doi.org/10.1109/ACCESS.2019.2932314
22. Yakubova R, Han S, Kim Y. Secure VoIP architecture using Asterisk PBX with TLS-enabled SIP channels: A practical implementation and Wireshark analysis. Journal of Communications and Networks. 2023; 25(2):165-177. Doi: https://doi.org/10.23919/JCN.2023.0009