OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

Voice over IP (VOIP) Security Research- A Research

Ashis Saklani[1], Rajni Nautiyal[2]
[1,] *Assistant Professor*, Dept. of CSE/MCA, *BTKIT(Kumaon Institute of Technology), Dwarahat, Almora*

ABSTRACT: *This is a work based on a survey of Voice over IP security research.Goal is to provide a roadmap for researchers seeking to understand existing capabilities and, and to identify gaps in addressing the numerous threats and vulnerabilities present in VoIP systems. We also briefly discuss the implications of our findings with respect to actual vulnerabilities reported in a variety VoIP products.*

I. INTRODUCTION

Voice over IP technologies are being increasingly adopted by consumers, enterprises, and telecoms operators due to their potential for higher flexibility, richer feature set, and reduced costs relative to their Public Switched Telephony Network (PSTN) coun- terparts. At their core, VoIP technologies enable the transmission of voice in any IP network, including the Internet. Because of the need to seamlessly interoperate with the existing telephony infrastructure, the new features, and the speed of development and deployment, VoIP protocols and products have been repeatedly found to contain numerous vulnerabilities [16] that have been exploited [19]. As a result, a fair amount of research has been directed towards addressing some of these issues. However, the effort is unbalanced, with little effort is spent on some highly deserving problem areas. We have conducted a comprehensive survey of VoIP security research, complement- ing our previous work that analyzed known vulnerabilities [16]. Our long-term goal is four-fold. First, to create a roadmap of existing work in securing VoIP, towards reducing the start-up effort required by other researchers to initiate research in this space. Second, to identify gaps in existing research, and to help inform the security community of challenges and opportunities for further work. Third, to provide an overall sanity check on the overall VoIP security research ecosystem, using known vulnerabilities as a form of ground truth. Finally, in the context of the VAMPIRE project [1] (which supported this work), to provide guidanceas to what further work in needed to better understand and analyze the activities of VoIP-system attackers. Naturally, such ambitious goals require significantly more space than is available in a single conference paper.

In this paper, we provide a representative sample of the research works we surveyed. We classify these works according to the class of threat they seek to address, using the VoIP Security Alliance (VoIPSA) [54] threat taxonomy. Although we were forced to omit a large number of related works (which we hope to present in a comprehensive form in due time), this survey should be a good starting point for anyone interested in conducting research on VoIP security. We also briefly discuss the implications of our findings with respect to actual vulnerabilities reported in a variety VoIP products. In the remainder of this paper, Section 2 gives a brief overview of SIP, one of the most popular VoIP technologies. Section 3 summarizes the threat model defined by the VoIP Security Alliance. We then present our survey of the research literature on VoIP security in Section 4, and discuss some of the implications in Section 5.

II. SIP OVERVIEW

SIP [42] is an application-layer protocol standardized by the Internet Engineering Task Force (IETF), and is designed to support the setup of bidirectional communication ses- sions including, but not limited to, VoIP calls. It is somewhat similar to HTTP, in that it is text-based, has a request-response structure, and uses a user authentication mech- anism based on the HTTP Digest Authentication. However, it is an inherently state- ful protocol that supports interaction with multiple network components (*e.g.,* PSTN bridges), and can operate over UDP, TCP, and SCTP. The main SIP entities are endpoints (softphones or physical devices), a proxy server, a registrar, a redirect server, and a location server. The registrar, proxy and redirect servers may be combined, or they may be independent entities. Endpoints communicate with a registrar to indicate their presence. This information is stored in the location server. A user may be registered via multiple endpoints simultaneously. During call setup, the endpoint communicates with the proxy, which uses the location server to de- termine where the call should be routed to. This may be another endpoint in the same network (*e.g.*, in the same enterprise), or another proxy server in another network. Alternatively, endpoints may use a redirect server to directly determine where a call should be directed to; redirect servers consult the location server in the same way that proxy servers operate during call setup. Once an end-to-end

channel has been established (through one or more proxies) between the two endpoints, SIP negotiates the session parameters (codecs, RTP ports, *etc.*) using the Session Description Protocol (SDP). In a two-party call setup between Alice and Bob, Alice sends an INVITE message to her proxy server, optionally containing session parameter information encoded within SDP. The proxy forwards this message directly to Bob, if Alice and Bob are users of the same domain. If Bob is registered in a different domain, the message will be relayed to Bob's proxy, and thence to Bob. The message may be forwarded to several endpoints, if Bob is registered from multiple locations. While the call is being set up, Alice is sent RINGING messages. Once the call has been accepted, an OK message is sent to Alice, containing Bob's preferred parameters encoded within SDP. Alice responds with an ACK message. Alice's session parameter preferences may be encoded in the INVITE or the ACK message. Following this exchange, the two endpoints can begin transmitting voice, video or other content using the agreed-upon media transport protocol, typically RTP. While the signaling traffic may be relayed through a number of SIP proxies, the media traffic is exchanged directly between the two endpoints. When bridging different networks, *e.g.,* PSTN and SIP, media gateways may disrupt the end-to-end nature of the media transfer to translate content between the formats supported by these networks. There are many other protocol interactions supported by SIP, that cover a number of common (and uncommon) scenarios including call forwarding (manual or automatic), conference calling, voicemail, *etc.* Typically, this is done by semantically overloading SIP messages such that they can play various roles in different parts of the call. SIP can use S/MIME to carry complex authentication payloads, including public key certificates. When TCP is used as the transport protocol, TLS can be used to protect the SIP messages. TLS is required for communication among proxies, registrars and redirect servers, but only recommended between endpoints and proxies or registrars. IPsec may also be used to protect all communications, regardless of transport protocol.

III. VOIP THREATS

As a starting point, we use the taxonomy provided by the Voice over IP Security Alliance (VoIPSA) [54]. VoIPSA is a vendor-neutral, not for profit organization composed of VoIP and security vendors, organizations and individuals with an interest in securing VoIP protocols, products and installations. In addition, we place the surveyed vulnerabilities within the traditional threat space of confidentiality, integrity, availability (CIA). Finally, we consider whether the vulnerabilities exploit bugs in the protocol, implementation or system configuration.

In future work, we hope to expand the number of views to the surveyed vulnerabilities and to provide more in-depth analysis. The VoIPSA security threat taxonomy defines the security threats against VoIP deployments, services, and end users. The key elements of this taxonomy are:

1. Social threats are aimed directly against humans. For example, misconfigurations, bugs or bad protocol interactions in VoIP systems may enable or facilitate attacks that misrepresent the identity of malicious parties to users. Such attacks may then act as stepping stones to further attacks such as phishing, theft of service, or unwanted contact (spam).

2. Eavesdropping, interception, and modification threats cover situations where an adversary can unlawfully and without authorization from the parties concerned listen in on the signaling (call setup) or the content of a VoIP session, and possibly modify aspects of that session while avoiding detection. Examples of such attacks include call re-routing and interception of unencrypted RTP sessions.

3. Denial of service threats have the potential to deny users access to VoIP services. This may be particularly problematic in the case of emergencies, or when a DoS attack affects all of a user's or organization's communication capabilities (*i.e.,* when all VoIP and data communications are multiplexed over the same network which can be targeted through a DoS attack). Such attacks may be VoIP-specific (exploiting flaws in the call setup or the implementation of services), or VoIP-agnostic (*e.g.,* generic traffic flooding attacks). They may also involve attacks with physical components (*e.g.,* physically disconnecting or severing a cable) or through computing or other infrastructures (*e.g.,* disabling the DNS server, or shutting down power).

4. Service abuse threats covers the improper use of VoIP services, especially (but not exclusively) in situations where such services are offered in a commercial setting. Examples of such threats include toll fraud and billing avoidance [51,52].

5. Physical access threats refer to inappropriate/unauthorized physical access to VoIP equipment, or to the physical layer of the network.

6. Interruption of services threats refer to non-intentional problems that may nonetheless cause VoIP services to become unusable or inaccessible. Examples of such threats include loss of power due to inclement weather, resource exhaustion due to over-subscription, and performance issues that degrade call quality.

IV. SURVEY OF VOIP SECURITY RESEARCH

In the this section, we classify various research papers across the first four elements of the VoIPSA taxonomy (the last two relate to physical and non-security issues). We also include a *cross-cutting* category, which includes work that covers multiple areas (*e.g.*, proposing a security architecture), and an *overviews* category that includes works that survey vulnerabilities, threats, and security mechanisms. We give an indication as to how many total pieces of related work (including those described in the text) could be classified in that category but were omitted due to space limitations. The works that are discussed offer a representative view of the type of research activity in these problem areas.

Overviews (36 items) Persky gives a very detailed description of several VoIP vulnerabilities [32]. A long discussion of threats and security solutions is given by Thermos and Takanen [53]. Cao and Malik [8] examine the vulnerabilities that arise from introducing VoIP technologies into the communications systems in critical infrastructure applications. They examine the usual threats and vulnerabilities, and discuss mitigation techniques. They conclude by providing some recommendations and best practices to operators of such systems. Butcher *et al.* [7] overview security issues and mechanisms for VoIP systems, focusing on security-oriented operational practices by VoIP providers and operators. Such practices include the separation of VoIP and data traffic by using VLANs and similar techniques, the use of integrity and authentication for configuration bootstrapping of VoIP devices, authentication of signaling via TLS or IPsec, and the use of media encryption. They briefly describe how two specific commercial systems implement such practices, and propose some directions for future research.

Adelsbach *et al.* [2] provide a comprehensive description of SIP and H.323, a list of threats across all networking layers, and various protection mechanisms. A similar analysis was published by the US National Institute of Standards and Technology (NIST) [20]. Anwar *et al.* [3] identify some areas where the NIST report remains incomplete: counter-intuitive results with respect to the relative performance of encryption and hash algorithms, the non-use of the standardized Mean Opinion Score to evaluate call quality, and the lack of anticipation of RTP-based denial of service. They then propose the use of design patterns to address the problems of secure traversal of firewalls and NAT boxes, detecting and mitigating DoS attacks in VoIP, and securing VoIP against eavesdropping. Seedorf [45] overviews the security challenges in peer-to-peer (P2P) SIP. Threats specific to P2P-SIP include subversion of the identity-mapping scheme (which is specific to the overlay network used as a substrate), attacks on the overlay network routing scheme, bootstrapping communications in the presence of malicious first-contact nodes, identity enforcement (Sybil attacks), traffic analysis and privacy violation by intermediate nodes, and free riding by nodes that refuse to route calls or otherwise participate in the protocol other than to obtain service for themselves (selfish behavior).

Addressing social threats (49 items) Niccolini [29] discusses the difficulties in protecting against IP telephony spam (SPIT) and overviews the various approaches for blocking such calls, identifying the technical and operational problems with each. Possible building blocks for SPIT prevention include black/whitelists combined with strong identity verification to provide a reliable CallerID system, referral-based systems among trusted SIP domains, pattern or anomaly detection techniques to discriminate SPIT based on training data, multi-level grey-listing of calls based on caller behavior (similar to throttling), computational puzzles and CAPTCHAs, explicit callee consent (a form of capability, required to actually place a call), content filtering on voicemail spam, callee feedback to indicate whether a call was SPIT or legitimate (typically com- bined with white/blacklisting, and requiring strong identity), changing one's SIP address as soon as SPIT messages arrive, requiring a monetary fee for the first contact, and legal action. Niccolini argues that none of these methods by itself is likely to succeed, promotes a modular and extensible approach to SPIT prevention, and presents a high-level architecture that was designed for use in a commercial SIP router. Mathieu *et al.* [27] describe SDRS, an anti-SPIT system that combines several of these detection schemes and takes into consideration user and operator preferences.

The SPIDER project (SPam over Internet telephony Detection sERvice) released a public project report [38] providing an overview of SPIT threats and the relevant European legal framework (both on an EU and national basis). The second public project report [25] focuses on SPIT detection and prevention, summarizing some of the work done in this space and defining criteria for evaluating the efficiency of anti- SPIT mechanisms. They then classify prior work according to fulfillment of these criteria, expanding on the relative strengths and weaknesses of each approach. The third public project report [37] builds on the previous two reports, describing an anti-SPIT architectural framework. Elements of this architecture include improved authentication,white/blacklisting, behavior analysis, the use of computational puzzles for challenge/response, reputation management, and audio content analysis.

P¨orschmann and Knospe [34] propose a SPIT detection mechanism based on apply-ing spectral analysis to the audio data of VoIP calls to create acoustic fingerprints. SPIT is identified by detecting several fingerprints across a large number of different calls.Schlegel *et al.* [44] describe a framework for preventing SPIT. They argue for a modular approach to identifying SPIT, using hints from both signaling and media transfer. The first stage of their system looks at information that is available prior to accepting the call, while the second stage interacts with a caller (possibly prior to passing on the call to the callee). The various components integrated in their system include white/blacklists, call statistics, IP/domain correlation, and Turing tests. Their system also allows for feedback from the callee to be integrated into the scoring mechanism, for use in screening future calls. The evaluation focuses on scalability, by measuring the response time to calls as call volumes increase.

Quittek *et al.* [35] propose the use of *hidden* Turing tests to identify SPIT callers.As a concrete approach, they leverage the interaction model in human conversation minimizes the amount of simultaneous ("double") talk by the participants, and the fact that there is a short pause at the beginning of an answered call, followed by a statement by the callee that initiates the conversation. By looking for signs of violation of such norms, it is possible to identify na¨ıve automated SPIT callers. The authors implement their scheme and integrated it with a VoIP firewall.

Dantu and Kolan [17] describe the Voice Spam Detector (VSD), a multi-stage SPIT filter based on trust, reputation, and feedback among the various filter stages. The pri- mary filter stages are call pattern and volume analysis, black and white lists of callers, per-caller behavior profile based on Bayesian classification and prior history, and reputation information from the callee's contacts and social network. They provide a formal model for trust and reputation in a voice network, based on intuitive human behavior. They evaluate their system in a laboratory experiment using a small number of real users and injected SPIT calls.

Kolan *et al.* [18] use traces of voice calls in a university environment to validate a mathematical model for computing the nuisance level of an incoming call, using feed-back from the receivers. The model is intended to be used in predicting SPIT calls in VoIP environments, and is based on the history of prior communications between the two parties involved, which includes explicit feedback from the receiver indicating that a call is unwanted (at a particular point in time).

Balasubramaniyan *et al.* [4] propose to use call duration and social network graphs to establish a measure of reputation for callers. Their intuition is that users whose call graph has a relatively small fan-out and whose call durations are relatively long are less likely to be spammers. Conversely, users who place a lot of very short calls are likely to be engaging in SPIT. Furthermore, spammers will receive few (if any) calls. Their system works both when the parties in a call have a social network link between them, and when such a link does not exist by assigning global reputation scores. Users that are mistakenly categorized as spammers are redirected to a Turing test, allowing them to complete the call if the answer correctly. In a simulation-based evaluation, the authors determine that their system can achieve a false negative rate of 10% and a false positive rate of 3%, even in the presence of large numbers of spammers.

Addressing eavesdropping, interception, and modification threats (34 items) Wang *et al.* [55] evaluate the resilience of three commercial VoIP services (AT&T, Vonage and Gizmo) against man-in-the-middle adversaries. They show that it is possible for an attacker to divert and redirect calls in the first two services by modifying the RTP endpoint information included in the SDP exchange (which is not protected by the SIP Digest Authentication), and to manipulate a user's call forwarding settings in the latter two systems. These vulnerabilities permit for large-scale voice pharming, where unsus- pecting users are directed to fake interactive voice response systems or human repre- sentatives. The authors argue for the need for TLS or IPsec protection of the signaling. Zhang *et al.* show that, by exploiting DNS and VoIP implementation vulnerabilities, it is possible for attackers to perform man-in-the-middle attacks even when they are not on the direct communication path of the parties involved. They demonstrate their attack against Vonage, requiring that the attacker only knows the phone number and the IP address of the target phone. Such attacks can be used to eavesdrop and hijack the victims' VoIP calls. The authors recommend that users and operators use signaling and media protection, conduct fuzzing and testing of VoIP implementations, and develop a lightweight VoIP intrusion detection system to be deployed on the VoIP phone. Salsano *et al.* give an overview of the various SIP security mechanisms (as of 2002), focusing particularly on the authentication component. They conduct an eval-uation of the processing costs of SIP calls that involve authentication, under different transport, authentication and encryption scenarios. They show that a call using TLS and authentication is 2.56 times more expensive than the simplest possible SIP configuration (UDP, no security). However, a fully-protected a call takes only 54% longer to complete than a configuration that is more representative than the basic one but still offers no security; the same fully-protected call and has the same processing cost if the transport is TCP without any encryption (TLS). Of the overhead, approximately 70% is attributed to message parsing and 30% to cryptographic processing. With the advent of

Datagram TLS (DTLS), it \ is possible that encryption and integrity for SIP can be had for all configurations (UDP or TCP) at no additional cost. A similar conclusion is reached by Bilien *et al.* [6], who study the overhead in SIP call setup latency when using end-to-end and hop-by-hop security mechanisms. They consider protocols such as MIKEY, S/MIME, SRTP, TLS, and IPsec, concluding that the overall penalty of using full-strength cryptography is low. Barbieri *et al.* [5] had found earlier that when using VoIP over IPsec, performance can drop by up to 63%; however, it is questionable whether these results still hold, given the use of hardware accelerators and the more efficient AES algorithm in IPsec. Rebahi *et al.* analyze the performance of RSA as used in SIP for authentication and identity management (via public-key certificates and digital signatures), and describe the use of Elliptic Curve DSA (ECDSA) within this context to improve performance. Using ECDSA, their prototype can handle from 2 to 8 times as many call setup requests per second, with the gap widening as key sizes increase.

Guo *et al.* [14] propose a new scheme for protecting voice content that provides strong confidentiality guarantees while allowing for graceful voice degradation in the presence of packet loss. They evaluate their scheme via simulation and micro-benchmarks. However, Li *et al.* [23] show that the scheme is insecure. Kuntze *et al.* [21] propose a mechanism for providing non-repudiation of voice content by using digital signatures. Seedorf also proposes the use of cryptographically generated SIP URIs to protect the integrity of content in P2P SIP. Specifically, he uses self-certifying SIP URIs that encode a public key (or, more compactly, the hash of a public key). The owner of the corresponding private key can then post signed location binding information on the peer-to-peer network (*e.g.*, Chord) that is used by call initiators to perform call routing.

Petraschek *et al.* examine the usability and security of ZRTP, a key agreement protocol based on the Diffie Hellman key exchange, designed for use in VoIP environments that lack pre-established secret keys among users or a public key infrastructure (PKI). ZRTP is intended to be used with SRTP, which performs the actual content en-cryption and transfer. Because of the lack of a solid basis for authentication, which makes active man-in-the-middle attacks easy to launch, ZRTP uses Short Authentication Strings (SAS) to allow two users to verbally confirm that they have established the same secret key. The verbal communication serves as a weak form of authentication at the human level. The authors identify a relay attack in ZRTP, wherein a man-in-the- middle adversary can influence the SAS read by two legitimate users with who he has established independent calls and ZRTP exchanges. The attacker can use one of the legitimate users as an oracle to pronounce the desired SAS string through a number of means, including social engineering. The authors point out that SAS does not offer any security in some communication scenarios with high security requirements, *e.g.*, a user calling (or being called by) their bank. The authors implement their attack and demonstrate it in a lab environment. Wright *et al.* apply machine learning techniques to determine the language spoken in a VoIP conversation, when a variable bit rate (VBR) voice codec is used based on the length of the encrypted voice frame. As a countermeasure, they propose the use of block ciphers for encrypting the voice. In follow-on work [57], they use profile Hidden Markov Models to identify specific phrases in the encrypted voice stream with a 50% average accuracy, rising to 90% for certain phrases.

Addressing denial of service threats (19 items) Rafique *et al.* analyze the ro- bustness and reliability of SIP servers under DoS attacks. They launch a number of synthesized attacks against four well-known SIP proxy servers (OpenSER, PartySIP, OpenSBC, and MjServer). Their results demonstrate the ease with which SIP servers can be overloaded with call requests, causing such performance metrics as Call Com-pletion Rate, Call Establishment Latency, Call Rejection Ration and Number of Retransmitted Requests to deteriorate rapidly as attack volume increases, sometimes with as few as 1,000 packets/second. As an extreme case of such attacks large volumes of IN- VITE messages can even cause certain implementations to crash. While documenting the susceptibility to such attacks, this work proposes no defense strategies or directions. Reynolds and Ghosal describe a multi-layer protection scheme against flood-based application- and transport-layer denial of service (DoS) attacks in VoIP. They use a combination of sensors located across the enterprise network, continuously estimating the deviation from the long-term average of the number of call setup requests and successfully completed handshakes. Similar techniques have been used in detecting TCP SYN flood attacks, with good results. The authors evaluate their scheme via simulation, considering several different types of DoS attacks and recovery models. Ormazabal *et al.* describe the design and implementation of a SIP-aware, rule- based application-layer firewall that can handle denial of service (and other) attacks in the signaling and media protocols. They use hardware acceleration for the rule matching component, allowing them to achieving filtering rates on the order of hundreds of transactions per second. The SIP-specific rules, combined with state validation of the endpoints, allow the firewall to open precisely the ports needed for only the local and remote addresses involved in a specific session, by decomposing and analyzing the con-tent and meaning of SIP signaling message headers. They experimentally evaluate and validate the behavior of their prototype with a distributed testbed involving synthetic benign and attack traffic generation. Larson *et al.* experimentally

analyzed the impact of distributed denial of service (DDoS) attacks on VoIP call quality. They also established the effectiveness of low-rate denial of service attacks that target specific vulnerabilities and implementa- tion artifacts to cause equipment crashes and reboots. They discuss some of the possible defenses against such attacks and describe Sprint's approach, which uses regional "cleaning centers" which divert suspected attack traffic to a centralized location with numerous screening and mitigation mechanisms available. They recommend that critical VoIP traffic stay on private networks, the use of general DDoS mechanisms as a front-line defense, VoIP-aware DDoS detection and mitigation mechanisms, traffic policing and rate-limiting mechanisms, the use of TCP for VoIP signaling, extended protocol compliance checking by VoIP network elements, and the use of authentication mechanisms where possible.Sengar *et al.* describe vFDS, an anomaly detection system that seeks to identify flooding denial of service attacks in VoIP. The approach taken is to measure abnormal variations in the relationships between related packet streams using the Hellinger distance, a measure of the deviation between two probability measures. Using synthetic attacks, they show that vFDS can detect flooding attacks that use SYN, SIP, or RTP packets within approximately 1 second of the commencement of an attack, with small impact on call setup latency and voice quality. Conner and Nahrstedt [9] describe a semantic-level attack that causes resource exhaustion on stateful SIP proxies by calling parties that (legitimately or in collusion) do not respond.

This attack does not require network flooding or other high traffic volume attacks, making it difficult to detect with simple, network-based heuristics used against other types of denial of service attacks. They propose a simple algorithm, called *Random EarlyTermination* (RET) for releasing reserved resources based on the current state of the proxy (overloaded or not) and the duration of each call's ringing. They implement and evaluate their proposed scheme on a SIP proxy running in a local testbed, showing that it reduces the number of benign call failures when under attack, without incurring measurable overheads when no attack is underway. Zhang *et al.* describe a denial of service attack wherein adversaries flood SIP servers with calls involving URIs with DNS names that do not exist. Servers attempting to resolve them will then have to wait until the request times out (either locally or at their DNS server), before they can continue processing the same or another call. This attack works against servers that perform synchronous DNS resolution and only maintain a limited number of execution threads. They experimentally show that as few as 1,000 messages per second can cause a well provisioned synchronous resolution server to exhibit very high call drops, while simple, single-threaded servers can be starved with even 1 message per second. As a countermeasure, they propose the use of non-blocking DNS caches, which they prototype and evaluate. Luo *et al.* experimentally evaluate the susceptibility of SIP to CPU-based denial of service attacks. They use an open-source SIP server in four attack scenarios: basic request flooding, spoofed-nonce flooding (wherein the target server is forced to validate the authenticator in a received message), adaptive-nonce flooding (where the nonce is refreshed periodically by obtaining a new one from the server), and adaptive- nonce flooding with IP spoofing. Their measurements show that these attacks can have a large impact on the quality of service provided by the servers. They propose several countermeasures to mitigate against such attacks, indicating that authentication by itself cannot solve the problem and that, in some circumstances, it can exacerbate its sever- ity. These mitigation mechanisms include lightweight authentication and whitelisting, proper choice of authentication parameters, and binding nonces to client IP addresses.

Addressing service abuse threats (8 items) Zhang *et al.* present a number of exploitable vulnerabilities in SIP that can manipulate billing records in a number of ways, showing their applicability against real commercial VoIP providers. Their focus is primarily on attacks that create billing inconsistencies, *e.g.*, customers being charged for service they did not receive, or over-charged for service received. Some of these attacks require a man-in-the-middle capability, while others only require some prior interaction with the target (*e.g.*, receiving a call from the victim SIP phone device).

Abdelnur *et al.* [1] use AVISPA to identify a protocol-level vulnerability in the way SIP handles authentication [50]. AVISPA is a model checker for validating security protocols and applications using a high-level protocol specification and security-goals language that gets compiler into an intermediate format that can be consumed by a number of lower-level checkers. The attack is possible with the SIP Digest Authentication, whereby an adversary can reuse another party's credentials to obtain unautho- rized access to SIP or PSTN services (such as calling a premium or international phone line). This attack is possible because authentication may be requested in response to an INVITE message at any time during a call, and the responder may issue an INVITE message during a call either automatically (because of timer expirations) or through a user action (*e.g.*, placing the caller on hold in order to do a call transfer). While the solution is simple, it requires changes possibly to all end-device SIP implementations. This work is part of a bigger effort to apply testing and fuzzing toward identifying vulnerabilities in SIP protocols, implementations, and deployed systems. It is worth noting that this work has resulted in a number of vulnerability disclosures in the Common Vulnerabilities and Exposures (CVE) database and elsewhere.

Cross-cutting efforts (51 items) Wieser *et al.* [56] extend the PROTOS testsuite with a SIP-specific analysis fuzzing module. They then test their system against a number of commercial SIP implementations, finding critical vulnerabilities in all of them.

Gupta and Shmatikov [15] formally analyze the security of the VoIP protocol stack, including SIP, SDP, ZRTP, MIKEY, SDES, and SRTP. Their analysis uncovers a num- ber of flaws, most of which derive from subtle inconsistencies in the assumptions made in designing the different protocols. These include a replay attack in SDES that com- pletely break content protection, a man-in-the-middle attack in ZRTP, and a (perhaps theoretical) weakness in the key derivation process used in MIKEY. They also show several minor weaknesses and vulnerabilities in all protocols that enable DoS attacks. Dantu *et al.* [12] describe a comprehensive VoIP security architecture, composed of components distributed across the media gateway controller, the proxy server(s), the IP PBX, and end-user equipment. These components explicitly exchange information toward better training of filters, and creating and maintaining white/blacklists. Implicit feedback is also provided through statistical analysis of interactions (*e.g.*, call frequency and duration). The architecture also provisions for a recovery mechanism that incorporates explicit feedback and quarantining.

Wu *et al.* design an intrusion detection system, called SCIDIVE, that is specific to VoIP environments. SCIDIVE aims to detect different classes of intrusions, can operate with different viewpoints (on clients, proxies, or servers), and takes into consideration both signaling (*i.e.*, SIP) and media-transfer protocols (*e.g.*, RTP). SCIDIVE's ability to correlate cross-protocol behavior, theoretically allows for detection of more complex attacks. However, the system is rules-based, which limits its effectiveness against new/unknown attacks. In follow-on work, Wu *et al.* [60] develop SPACEDIVE, a VoIP-specific intrusion detection system that allows for correlation of events among distributed rules-based detectors.

They demonstrate the ability of SPACEDIVE to detect certain classes of attacks using a simple SIP environment with two domains, and compare it with SCIDIVE. Niccolini *et al.* design an intrusion detection/intrusion prevention system archi- tecture for use with SIP. Their system uses both knowledge-based and behavior-based detection, arranged as a series in that order. They develop a prototype implementation using the open-source Snort IDS. They evaluate the effectiveness of their system in an attack scenario by measuring the mean end-to-end delay of legitimate SIP traffic in the presence of increasing volumes of malformed SIP INVITE messages. Nassar *et al.* advocate the use of SIP-specific honeypots to catch attacks targeting the Internet telephony systems, protocols and applications. They design and implement such a honeypot system, and explore the use of a statistical engine for identifying attacks and other misbehavior, based on training on legitimate traces of SIP traffic. The engine is based on their prior work that uses Bayesian-based inference. The resulting SIP honeypot effort is largely exploratory, with performance and effectiveness evaluations left for future work.

Rieck *et al.* apply machine learning techniques to detecting anomalous SIP messages, incorporating a "self-learning" component by allowing for periodic re-training of the anomaly detector using traffic that has been flagged as normal. The features used for clustering are based on n-grams and on tokenization of theSIP protocol. To prevent training attacks, wherein an adversary "trains" the anomaly detector to accept malicious inputs are legitimate, they employ randomization (choosing random samples for the training set),sanitization [10], and verification (by comparing the output of the new and old training models). Their experimental prototype was shown to handle 70 Mbps of SIP traffic, while providing a 99% detection rate with no false positives. SNOCER, a project funded by the European Union, is "investigating approaches for overcoming temporal network, hardware and software failures and ensuring the high availability of the offered VoIP services based on low cost distributed concepts." The first public project report [48] provides an overview of VoIP infrastructure components and the threats that must be addressed (staying primarily at the protocol and network level, and avoiding implementation issues with the exception of SQL injection), along with possible defense mechanisms.

There is also discussion on scalable service provisioning (replication, redundancy, backups *etc.*), toward providing reliability and fault tolerance. The second public project report [11] describes an architecture for protecting against malformed messages and related attacks using specification-based intrusion detection, protocol message verification, and redundancy. They use ontologies to describe SIP vulnerabilities, to allow for easy updating of the monitoring components (IDS) [13]. Marshall *et al.* describe the AT&T VoIP security architecture. They divide VoIP equipment into three classes: trusted, trusted-but-vulnerable, and untrusted. The latter consists of the customer premises equipment, which is outside the control of the carrier. The trusted domain includes all the servers necessary to provide VoIP service. Between the two sit various border and security elements, that are responsible for protecting the trusted devices while permitting legitimate communications to proceed. They describe the interactions among the various components, and the security mechanisms used in protecting these interactions.

V. DISCUSSION

In our previous work [16], we surveyed over 200 vulnerabilities in SIP implementations that had been disclosed in the CVE database from 1999 to 2009. We classified these vulnerabilities along several dimensions, including the VoIPSA threat taxonomy, the traditional Confidentiality, Integrity, Availability concerns, and a Protocol, Implementation, Configuration axis. We found that the various types of denial of service attacks constitute the majority of disclosed vulnerabilities, over 90% of which were due to implementation problems and 7% due to configuration. Considering the research work we have surveyed (some of which was discussed in this paper), we can see that out of a total of 197 publications, 18% concern themselves with an overview of the problem space and of solutions — a figure we believe is reasonable, considering the enormity of the problem space and the speed of change in the protocols, standards, and implementations. We also see a considerable amount of effort (roughly 25%) going toward addressing SPIT. While SPIT is not a major issue at this point, our experience with email spam and telemarketing seems to provide sufficient motivation for research in this area. Much of the work is focused on identifying SPIT calls and callers based on behavioral traits, although a number of other approaches are under exploration (*e.g.*, real-time content analysis). One of the problems is the lack of a good corpus of data for experimentation and validation of the proposed techniques.We were also not surprised to see a sizable portion of research (17%) directed at design, analysis (both security- and performance-oriented), and attacking of cryptographic protocols as used in VoIP. The cryptographic research community appears to be reasonably comfortable in proposing tweaks and minor improvements to the basic au-thentication mechanisms, and the systems community appears content with analyzing the performance of different protocol configurations (*e.g.*, TLS *vs.* IPsec). With a few notable exceptions, much of the work lacks "ambition."

Most distressing, however, is the fact that comparatively little research (9.6%) is go- ing toward addressing the problem of denial of service. Given the numerical dominance of SIP-specific DoS vulnerabilities (as described earlier) and the ease of launching such attacks, it is clear that significantly more work is needed here. What work is being done seems to primarily focus on the server and infrastructure side, despite our finding that half of DoS-related vulnerabilities are present on endpoints. Furthermore, much of the existing work focuses on network-observable attacks (*e.g.*, "obviously" malformed SIP messages), whereas the majority of VoIP DoS vulnerabilities are the result of implemen- tation failures. More generally, additional work is needed in strengthening implemen- tations, rather than introducing middleboxes and network intrusion detection systems, whose effectiveness has been shown to be limited in other domains; taking a black box approach in securing VoIP systems is, in our opinion, not going to be sufficient. Also disconcerting is the lack of research (4%) in addressing service abuse threats, considering the high visibility of large fraud incidents [19,51,52]. In general, we found little work that took a "big picture" view of the VoIP security problem. What cross cutting architectures have been proposed focus primarily on intrusion detection. Work is desperately needed to address cross-implementation and cross-protocol problems, above and beyond the few efforts along those lines in the intrusion detection space. Finally, we note that none of the surveyed works addressed the problem of configu- ration management. While such problems represent only 7% of known vulnerabilities, configuration issues are easy to overlook and are likely under-represented in our previ- ous analysis due to the nature of vulnerability reporting.

VI. CONCLUSIONS

We have presented a survey of VoIP security research. While space restrictions prevent us from discussing all surveyed works, we have discussed a representative subset of these. We presented an initial classification using the VoIPSA threat taxonomy, and juxtaposed this against our previous analysis on VoIP security vulnerabilities. We iden- tified two specific areas (denial of service and service abuse) as being under-represented in terms of research efforts directed at them (relative to their importance in the vulnerability survey), and called for additional effort at securing implementations and configurations, rather than taking a black-box approach of VoIP systems. We intend to expand on this work and offer a more comprehensive analysis in the near future.

REFERENCES

[1]. H. Abdelnur, T. Avanesov, M. Rusinowitch, and R. State. Abusing SIP Authentication. In Proceedings of the 4th International Conference on Information Assurance and Security (ISIAS, pages 237-242, September 2008.

[2]. A. Adelsbach, A. Alkassar, K.-H. Garbe, M. Luzaic, M. Manulis, E. Scherer, J. Schwenk, and E. Siemens. Voice over IP: Sichere Umstellung der Sprachkommunikation auf IP- Technologie. Bundesanzeiger Verlag, 2005.

[3]. Z. Anwar, W. Yurcik, R. E. Johnson, M. Hafiz, and R. H. Campbell. Multiple Design Patterns for Voice over IP (VoIP) Security.

[4]. In Proceedings of the IEEE Workshop on Informational Assurance (WIA), held in conjunction with the 25th IEEE International Performance Com- puting and Communications Conference, (IPCCC), April 2006.

[5]. V. Balasubramaniyan, M. Ahamad, and H. Park. CallRank: Combating SPIT Using Call Duration, Social Networks and Global Reputation. In Proceedings of the 4th Email and Anti-Spam (CEAS), August 2007.

[6]. R. Barbieri, D. Bruschi, and E. Rosti. Voice over IPsec: Analysis and Solutions. In Proceedings of the 18th Annual Computer Security Applications Conference (ACSAC), pages 261-270, December 2002.

[7]. J. Bilien, E. Eliasson, J. Orrblad, and J.-O. Vatn. Secure VoIP: Call Establishment and Media Protection. In Proceedings of the 2nd Workshop on Securing Voice over IP, June 2005.

[8]. D. Butcher, X. Li, and J. Guo. Security Challenge and Defense in VoIP Infrastructures. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 37(6):1152-1162, November 2007.

[9]. F. Cao and S. Malik. Vulnerability Analysis and Best Practices for Adopting IP Telephony in Critical Infrastructure Sectors. IEEE Communications Magazine, 44(4):138-145, April 2006.

[10]. W. Conner and K. Nahrstedt. Protecting SIP Proxy Servers from Ringing-based Denial-of- Service Attacks. In Proceedings of the 10th (ISM), pages 340-347, December 2008.

[11]. G. F. Cretu, A. Stavrou, M. E. Locasto, S. J. Stolfo, and A. D. Keromytis. Casting out Demons: Sanitizing Training Data for Anomaly Sensors. In Proceedings of the IEEE Security and Privacy Symposium, pages 81-95, May 2008.

[12]. T. Dagiuklas, D. Geneiatakis, G. Kambourakis, D. Sisalem, S. Ehlert, J. Fiedler, J. Markl, M. Rokis, O. Botron, J. Rodriguez, and J. Liu. General Reliability and Security Framework for VoIP Infrastructures. Technical Report Deliverable D2.2, SNOCER COOP-005892, September 2005.

[13]. R. Dantu, S. Fahmy, H. Schulzrinne, and J. Cangussu. Issues and Challenges in Securing VoIP. Computers & Security (to appear), 2009.

[14]. D. Geneiatakis and C. Lambrinoudakis. An Ontology Description for SIP Security Flaws.Computer Communications, 30(6):1367-1374, April 2007.

[15]. J.-I. Guo, J.-C. Yen, and H.-F. Pai. New Voice over Internet Protocol Technique with Hierarchical Data Security Protection. IEE Proceedings — Vision, Image and Signal Processing, 149(4):237-243, August 2002.

[16]. P. Gupta and V. Shmatikov. Security Analysis of Voice-over-IP Protocols. In Proceedings of the 20th 2007. IEEE Computer Security Foundations Symposium (CSFW), pages 49-63, July D. Keromytis. Voice over IP: Risks, Threats and Vulnerabilities. In Proceedings of the Cyber Infrastructure Protection (CIP) Conference, June 2009.

[17]. P. Kolan and R. Dantu. Socio-technical Defense Against Voice Spamming. ACM Transactions on Autonomous and Adaptive Systems (TAAS), 2(1), March 2007.

[18]. P. Kolan, R. Dantu, and J. W. Cangussu. Nuisance of a Voice Call. ACM Transactions on Multimedia Computing, Communications and Applications (TOMCCAP), 5(1):6:1-6:22, October 2008.

[19]. B. Krebs. Security Fix: Default Passwords Led to $55 Million in Bogus Phone Charges, June 2009.

[20]. D. R. Kuhn, T. J. Walsh, and S. Fries. Security Considerations for Voice Over IP Systems.US National Institute of Standards and Technology (NIST) Special Publication SP 800-58, January 2005.