SDRS: A Voice-over-IP Spam Detection and Reaction System

An expected surge in spam over Internet telephony (SPIT) requires a solution that incorporates multiple detection methods and reaction mechanisms, enabling greater flexibility and customization.

In general, “spam” describes information, often dubious in nature, sent to numerous recipients without their prior consent. Although the term typically refers to emails about hot stocks, revolutionary medicine, or adult content, spam can apply to all kinds of messages. Examples range from telemarketing calls and short message service texts to bulk mail and faxes.

Since the first incident in the early ’90s, Internet spam has increased significantly. Of all exchanged mail, spam’s portion has risen from less than 10 percent in 2001 to more than 80 percent today, according to statistics from antispam organizations such as Spam-O-Meter.com.

Session Initiation Protocol (SIP) has established itself as the de facto standard for voice-over IP (VoIP) services in fixed and mobile environments. From a technological viewpoint, SIP-based VoIP services show a greater resemblance to email than to traditional telephony systems. Hence, with SIP services gaining in popularity, spammers likely will misuse services as they do email—a practice known as spam over Internet telephony (SPIT).

This probable exponential increase in spam requires mitigating SPIT in its early stages. Solutions are even more critical because of SPIT’s threat to users’ trust in VoIP in general. Lack of confidence in secure and trusted infrastructures would slow down VoIP adoption.

Our solution framework combines well-known detection schemes, such as blacklists and white lists, with methods based on statistical traffic analysis, such as the number and duration of calls a user conducts.

Email vs. VoIP Spam

Why the expected surge in SPIT? Compared with email, using voice calls offers spammers a wider range of use scenarios:

- **Passive marketing.** Most spam email offers fall into this category. With SPIT, a prerecorded voice or voice/video message presents the sales pitch. Once a recipient accepts a call, the system delivers the content as a media stream.
- **Interactive marketing.** These are the standard telemarketing calls in which a live caller tries to sell goods or services, such as insurance or financial services, to a callee.
- **Call back.** In this method of fraud common to mobile networks, the fraudster calls a mobile phone number but hangs up just before the callee answers. Out of curiosity, the callee returns the call, unaware that it’s a premium phone number, and incurs a hefty charge.

Although spammers can conduct these types of unsolicited calls using traditional public switched telephone network (PSTN) telephony services, SIP offers advantages in cost, scope, identity hiding, and regulation.

The difference between per-minute costs for VoIP and PSTN is vanishing in some countries, such as Ger-
many and the US, but the cost gap in many regions is still significant. That said, even with equal per-minute prices, a spammer using a flat-rate DSL line with 2 Mbits can conduct about 30 calls in parallel using VoIP. For PSTN, this level of efficiency would generally require more expensive PSTN lines. Furthermore, SPIT can use software running on off-the-shelf PCs, whereas PSTN requires special hardware.

Cost benefits also affect scope. Launching an international spam campaign using PSTN can be rather expensive. With VoIP, however, a spammer can subscribe to a flat-rate service abroad at costs similar to a national campaign. A spammer in Germany, for example, can use a US VoIP service.

SIP-based services also make it easier to conceal a spammer’s identity. Using open-relay servers and forging headers makes identity hiding in email rather simple. High bandwidth usage when sending millions of messages over the same Internet access would probably alert the ISP to possible misuse. However, to avoid arousing suspicion, a spammer can distribute the load by using a botnet comprising thousands of bots. With spam over PSTN, disguising this high call volume is more difficult. Although the spammer can anonymize calls, the operator can trace the calls’ origin and block them. In this regard, SIP-based services are more like email services. The spammer can use botnets for distributing the load and relay the SIP request over SIP proxies, which relay requests without authenticating them first.

If a spammer’s identity is discovered, VoIP offers a way around government regulations and steep penalties. Although the number of telemarketing calls that can be classified as spam is still relatively small compared with email spam, most countries have some regulations outlawing unsolicited marketing calls. Spammers would be subject to such regulations and risk heavy penalties. Due to the high costs of international calls, telemarketers usually restrict their activities to their domestic markets. By reducing the costs for conducting spam abroad, VoIP lets spammers work across borders, thus falling outside the jurisdiction of such regulations.

The advantages that VoIP has over email for spammers could result in an alarming increase in SPIT levels, hence the need for our solution.

**VoIP Traffic Characteristics**

One of the approaches our SDRS solution uses to detect SPIT is based on identifying anomalies in the number or duration of calls a user conducts and the percentage of failed calls. We first classified normal user behavior, such as that of nonspammers. To do so, we monitored and analyzed the VoIP traffic of 8,700 France Telecom VoIP users, capturing usage in nine locations in France over a one-month period.

![Figure 1. Number of calls during a one-month period. An excessive number of calls from one user can indicate a possible spitter.](image-url)

The average number of calls per user was about two and a half per working day (Monday through Friday) and about two per weekend day. The maximum number of calls for a single user was about 20 per day, a figure still small enough to fall into the normal range for users such as teenagers. For a telemarketing service, the number of calls would be higher. System administrators could therefore specify a maximum acceptable value for the number of calls (for example, more than 30) before classifying the caller as a spitter. Furthermore, the average call duration is roughly six minutes per call. So, administrators could classify frequent calls that are, for example, only one minute long as SPIT.

Figure 1 shows a distribution of the number of calls per user over one month. Although the maximum number of calls by one user during the entire month is about 500, half of the users initiate fewer than 40 calls per month (or just over one a day). Figure 2 shows the distribution of the total duration of user calls during the same month. Half of the users called for about three hours (or 10,000 seconds) per month. The maximum was roughly 30 hours per month.

Finally, checking call destination, we observed that more than 90 percent of the calls that French users initiated stayed within France, as Figure 3 shows. Numerous calls from a single user to a foreign country could indicate a spitter.

Although these results are valuable in fine-tuning a detection system, they’re most likely only of limited scope. Traffic characteristics for different networks will depend on customer type (such as enterprise or private), age, habits, and nationality, as well as the
structure of local PSTN minute prices. So, the first step in introducing a SPIT detection system should be thoroughly analyzing the VoIP traffic.

**SPIT Detection and Reaction System**

As with spam, each approach for preventing SPIT has its advantages and disadvantages. Detecting and mitigating SPIT with only one of these methodologies would be impossible. Considering SPIT’s various forms, different user preferences, and spam’s ever-changing characteristics, an efficient anti-SPIT solution should incorporate several detection methods weighted according to user preferences and adapted to traffic characteristics.

As Figure 4 illustrates, SDRS comprises three main parts: user classification, SPIT detection, and reaction. By monitoring users’ sending behavior, SDRS classifies users as malicious (such as spitters) or normal. SDRS also collects recipient preferences. The system then uses this data as input for the detection and reaction components. Next, SDRS combines detection methods that enable the system to classify a call as SPIT both during session establishment and after the call has been established. Finally, SDRS implements various reactions to SPIT calls, such as rejecting suspicious calls and limiting suspected spitters’ call-generation rate.

To align SDRS’s capacity with increased VoIP and SPIT volume, we developed SDRS as a distributed system. The detection and reaction components can be either collocated or implemented on different hosts. The SPIT reaction system then can be collocated with the detection systems. Alternatively, it can be located remotely in a centralized server serving other multiple detection systems (for example, for investment reduction) or in a peripheral part of the network to accommodate user-specific preferences (for example, the user terminal).

Distributing the detection and reaction components improves the system’s scalability and flexibility. However, it also can increase the complexity by requiring additional management components for coordinating detection and reaction. To avoid this, we integrate the coordination information as part of the SIP messages themselves. More specifically, if an SDRS Detection System component identifies a call as SPIT, the system marks SIP messages with a new field that indicates a SPIT-level score. The reaction component can then adjust its actions according to this information.

**User Classification**

Various conventions on human rights and freedom indicate that intercepting, opening, reading, or delaying reception of communications or impeding message sending intrudes on the right of correspondence. From a sender’s viewpoint, message blocking limits freedom of speech. From a recipient’s viewpoint, it amounts to censorship.

For a service provider to protect its infrastructure, reputation, and customers, while avoiding intruding on subscribers’ privacy, it must put subscribers in charge of the anti-SPIT measures. To accomplish this, service providers can offer subscribers a cost-free way...
to opt-out of protective measures. The user classification component is a database that collects recipient preferences for which calls they want to receive.

One detection mechanism SDRS employs is based on blacklists. Calls from blacklisted users have a high probability of getting rejected. To adjust blacklists dynamically, SDRS classifies callers as frequent, occasional, or nonspitter on the basis of the number of SPIT calls they send. The system adjusts the reaction according to the classification category. To accommodate changes in the user’s behavior, the classification mechanisms can move users from one category to another, depending on the call history and the associated SPIT level.

This classification is also useful for detecting changes in user behavior. Indeed, when users who have never or rarely initiated SPIT suddenly send a large amount of SPIT calls, the abnormal activity might indicate a spoofing attack or virus. This classification can help detect potential virus propagation on previously safe devices and initiate or suggest that the service or network provider perform a virus check for users.

**SPIT Detection System**

The characteristic that distinguishes our detection system is that it addresses the SPIT threat with a modular and extensible architecture, bringing together many detection methods for SPIT identification. It enables online addition, update, and configuration of modules for quick and automatic reaction to changes in the SPIT threat.

As Figure 5 shows, the SPIT detection system comprises detection modules that take as input SIP messages (such as INVITE or BYE) and process them to detect SPIT. Each module outputs the likelihood that a certain SIP message is SPIT. The multiple detection methods combine for a scoring system that weights and totals the SPIT likelihood scores and computes the system’s estimation of each message’s SPIT level. The category of the user initiating the call influences the detection modules used and their weights. The SPIT level therefore computes as

\[
\text{spitLevel} = \left( \sum w_i \cdot \text{spitLikelihood}_i \right)_{\text{user classification dependent}}
\]

The higher the SPIT level the more likely the message is SPIT.

In general, a perfect anti-SPIT solution would detect and prevent all SPIT calls before they reach the recipient. If this online detection step fails, the SPIT call reaches the user. To reduce the possibility of such failures in the future, SDRS also uses offline detection. The system takes information it collects about calls identified as SPIT only after receipt and uses the data to refine the online detection step and detect SPIT for the next calls.

**Detection before call establishment (online).** These methods aim to detect SPIT calls before call establishment, while requiring the least possible interaction with caller and callee. Therefore, we don’t consider techniques that involve such interaction, such as the audio Completely Automated Public Turing to tell Computer from Humans Apart (CAPTCHA) or Voice Printing. SDRS uses four primary detection criteria for computing SPIT likelihood scores before call establishment: white lists, blacklists, caller behavior monitoring, and spoofing detection.
A white list module compares the caller identity to a database of stored trusted identities. The system administrator can manually set a white list to indicate callers the operator trusts. Furthermore, these lists also consider the recipients’ preferences for which callers they do and don’t trust.

A blacklist module compares the caller identity to database identities deemed not to be trusted. Administrators can manually set up blacklists and base them on classification results. For example, the system dynamically adds to the blacklist those callers it has detected have sent numerous SPIT calls.

In caller behavior, monitoring detects deviations to the expected normal behavior. The system monitors three parameters:

- **Simultaneous calls.** This module checks if the caller has multiple simultaneous active calls. If the number of parallel calls a user initiates is above a certain threshold, the probability that the user is a spitter is high.
- **Call rate.** This module checks if the caller has been placing numerous calls during the last period of time. A large number of calls is a good indication of SPIT activity.
- **Number of error messages associated with the caller.** This module checks if the caller has been receiving a lot of error messages in answer to its session initiation messages (corresponding to a nonexistent callee) during the last time period. Spitters often use lists of user names generated from a dictionary. For example, take a well-known domain name and try variations on that name. Given that many of the generated names don’t exist, a large number of calls will fail. Hence, a large call failure rate can indicate SPIT activity as well.

A major drawback of using black and white lists as well as traffic analysis based on monitoring caller behavior is that these mechanisms require using authenticated identities. Although standardization bodies are currently working on this requirement, authenticated identities’ usage isn’t widely spread. Therefore, in this part of the detection process, SDRS aims at detecting spoofing attempts through pattern detection in callee timing, IP and domain correlation, and statistical analysis.

- **Pattern detection in callee dialing.** This module checks for a deterministic pattern associated with callee names (for example, the system receives calls in sequence for bob@biloxi.com, boba@biloxi.com, bobb@biloxi.com, and so on).
- **IP and domain correlation.** This module checks for a correlation between IP addresses, user names, and domains of the user initiating sessions in a certain time period. SDRS can use this information to detect attackers who are spoofing their SIP addresses to trick the detection system. So, for example, it’s rather suspicious if multiple users are using the same IP address, even more if these multiple users are using various domain names. Similarly, it’s also suspicious when a single user uses multiple IP addresses in a short time.
- **Statistical analysis.** This module checks for a statistical pattern behind session initiation messages. In particular, it checks statistical patterns between call initiations and call terminations to see if a machine is trying to place calls sequentially or following a predefined temporal distribution.

### Detection after call establishment (offline)

To improve detection and user classification efficiency, SDRS also collects information about calls identified as SPIT only after receipt. As part of offline detection, SDRS monitors call duration. This test consists of classifying a call as SPIT based on the duration of initiated calls.

Indeed, if a caller always plays the same message (like an automatic machine), the majority of the calls will have the same duration. The user classification component can then use this information. Users that generate numerous calls of the same length will move from one category to another in increasing order of maliciousness, depending on the number of observations. This check could also help sort the users’ mailboxes by moving messages into a folder labeled “SPIT,” making it easier to quickly identify SPIT calls rather than losing time checking each message’s content. If the callee is annoyed with the call’s content and terminates the session before the end of the message, clearly the call won’t match the default length. In this case, the call won’t contribute to the SPIT statistics and won’t help in refining the detection mechanisms.

To help with detection after call establishment, SDRS also implements feedback mechanisms as Saverio Niccolini and his colleagues suggested.³

### SPIT Reaction System (Online)

After detecting a SPIT call, the system should take action on the basis of the call’s SPIT level, the provider’s policies, and the user’s category, as Figure 4 depicts. Possible reactions include limiting the number of calls, temporary blacklisting, call redirection, and notification.

Limiting of the number of calls is intended for occasional spitters and involves fixing a maximum threshold of calls per unit of time using a sliding window (for example, three calls per hour). The system blocks calls generated in excess of this threshold. The system administrator can further adjust this reaction on the basis of the user’s category by defining different thresholds and time units.
Because of its similarity to spam, the obvious approach for fighting spam over Internet telephony (SPIT) is to deploy common antispam methods. These techniques include:

- **White lists.** Only calls from known users are allowed.
- **Blacklists.** Calls from blacklisted users are rejected.
- **Content analysis.** This is one of the most widely used spam detection schemes. However, analyzing audio or visual content requires more resources than analyzing textual messages.

Furthermore, this method is useful only for calls directed to a voicemail box. Once a user accepts a call, it's too late to deploy this approach, as the user can do the analysis at that point.

Besides these common approaches, many experimental solutions address SPIT. Yacine Rebahi and Dorgham Sisalem use white lists for building a reputation system that helps to classify users as trusted or not trusted. Marcus Hansen and colleagues describe a user-controlled system built on white and blacklists combined with a peer-to-peer Web of trust. In Donwook Shin and Choon Shim's work, a technique called progressive multi gray-leveling assigns a gray level to each caller based on the number of calls he or she places both on a short and on a long term. Hong Yan and colleagues introduce a SPIT firewall that uses fingerprinting for identifying the calling devices. Lately, work has also focused on usage of limited-use addresses and voice printing technology, as well as on detection using human communication pattern matching.

To reduce the possibility of misusing SIP for SPIT, various proposals have surfaced at standardization bodies. Souhwan Jung and his colleagues suggest an approach for authentication between inbound proxy and caller to prevent SPIT calls from trying to bypass a server's security checks. Saverio Niccolini and his colleagues propose SIP extensions that would enable SIP servers to collaborate in preventing SPIT by rating single messages as possibly SPIT (or not) and exchanging this information.

**References**


Temporary blacklisting is intended for frequent spitters. With this approach, the operator refuses all calls from the identified spitter for a certain time period.

Call redirection involves redirecting suspicious calls transparently toward an automat, for example, a voicemail box. The administrator then analyzes the recorded calls offline or uses them as proof of malicious behavior. Furthermore, this reaction can help in detecting potential virus propagation and initiating or suggesting a virus check to end-users.

In notification, the system notifies a network management and monitoring system about a call’s SPIT level, but the session initiation itself isn’t affected. The system then uses the information for user classification or general monitoring.

Because purposefully dropping or blocking calls generally isn’t allowed, operators must consider legal restrictions. For example, the service provider must inform its subscribers about the measures taken for detecting and blocking unsolicited communication, and the provider’s customers must agree to these measures. The provider should also offer its subscribers a cost-free way to opt-out of such protective measures. Moreover, the subscriber must have a cost-free way of viewing messages the provider has classified as SPIT. So, the subscriber can make the final decision about whether to delete a message. This approach reduces the possibility of a false positive, or a message wrongly classified as SPIT. Finally, the provider must clearly describe its policy for how it deals with collected data, such as spam samples and receiver and sender addresses, and under what terms and for what purposes the provider can exchange such data with other providers.
We’ve partially taken care of these considerations in SDRS’s design by incorporating user preferences in the system’s detection and reaction processes.

**Network Architecture**

For an anti-SPIT system to be able to fulfill its purpose, it must be able to monitor and collect the SIP messages. Therefore, placement in the network must be where the systems can access the signaling messages and mainly depends on the targeted architectures and operator’s deployment requirements. The possibilities include coupling the SPIT detection and reaction systems with servers already involved in signaling tasks, such as SIP proxy servers, session border controllers, and application servers. Alternatively, they can be coupled with additional elements located in the network (in other words, in the end-to-end path) and not specifically dedicated to such tasks, such as deep packet inspection (DPI) devices. Locating the system in the user device itself is also a possibility, but we don’t discuss it in this article.

**Signaling Servers**

Signaling elements, such as SIP proxy servers or session border controllers, process all requests targeted to the service provider’s users and, thus, constitute an optimal placement for SPIT detection and reaction systems.

When integrating an anti-SPIT system with a signaling component, an integration effort is needed to bring the detection and reaction logic into the server itself (specifically, it must have access to management APIs). Integration should be done carefully to avoid overloading the VoIP servers, already busy with call establishment.

An alternative solution for reducing the integration effort is to insert an additional specific element in the signaling path to separate the detection and reaction systems from the signaling components. Such a scenario would use dedicated SPIT detection and reaction application servers. Although such an approach would simplify the introduction of anti-SPIT systems, it adds another component to the system, increasing the complexity and the number of possible failure points of the VoIP infrastructure. It further increases the SIP messages’ processing delay.

**DPI Devices**

DPI devices are located in the network and are able to analyze real-time traffic up to the application layer (Open Systems Interconnection layer 7). Currently, DPIs are mainly employed in bandwidth management, like traffic limitation or traffic redirection for Web caches. Here, we consider extending DPIs to manage VoIP calls to detect SPIT calls and eventually to carry out adequate reactions if needed. In this case, the network operator could deploy the anti-SPIT solutions at different levels of an IP network: at point of presence, at the broadband access server for the Asymmetric Digital Subscriber Line access network, close to access points for the WiFi network, or at the peering points between different providers, as Figure 6 shows.

Unlike the integration with signaling servers, the main advantage of such an approach is that administrators can use the architecture for detecting SPIT in centralized VoIP infrastructures (comprising SIP proxies and SBCs) as well as in distributed peer-to-peer (P2P) architectures. This point is especially interesting considering the recent advances in standardization of SIP-based P2P technology.7

Another advantage of this solution is that it can check the control and the data plane simultaneously, thereby detecting SPIT both on the signaling as well as the media level. With the expected increase in programmability, extensibility, and performance, DPI devices will one day be able to analyze in real time voice traffic in a scalable manner. Such DPIs would then be able to detect the same vocal message played several times or to look for known commercial sentences in the media stream.

Moreover, deploying such a solution in network equipment located in the caller’s access network will enable rapid detection of SPIT at the entry of the network before overloading the SIP servers, the network, and eventually the users’ devices.

Finally, perhaps out of scope for this article, we can also use this kind of architecture and DPI equipment for other traffic monitoring, such as fraud and denial-of-service attacks. From a deployment perspective, this can be an advantage for network operators willing...

---

**Figure 6. Deep packet inspection placement in a network architecture.** Administrators can deploy anti-SPIT solutions at different levels of an IP network.
to reuse investments. On the negative side, managing numerous detection points increases the management overhead for keeping the detection and reaction rules up-to-date.

On one hand, the flexibility and low cost structure of VoIP is rapidly attracting new customers and enabling innovative services. On the other hand, VoIP provides a powerful and flexible tool for conducting unsolicited communication, for example, SPIT.

Our work provides the framework for an anti-SPIT solution. As part of our future work, we will extend the system in different aspects. Based on additional real live traffic traces, we’ll establish more general models for user behavior. We’ll also integrate additional detection mechanisms on the basis of reputation to enhance the white list and blacklist mechanisms. From a deployment perspective, the parameters the system uses, such as reaction thresholds and classification categories, must adapt to real requirements and user and operator preferences to limit the false positives. Finally, from a technical perspective, we’ll have to assess the effects of deploying an anti-SPIT solution on processing and memory requirements.

References


Bertrand Mathieu is a senior researcher at Orange Labs. His research interests include programmable networks, adaptive application-level network solutions, and peer-to-peer networks. Mathieu has a PhD from University Pierre et Marie Curie (LIP6) of Paris. Contact him at bertrand2.mathieu@orange-ftgroup.com.

Saverio Niccolini is a manager at NEC Laboratories Europe. His research interests include VoIP security, monitoring, and peer-to-peer applications. Niccolini has a PhD from the University of Pisa. Contact him at saverio.niccolini@nw.neclab.eu.

Dorgham Sisalem is the director for Strategic Architecture at Tekelec. His research interests include VoIP security and congestion control. Sisalem has a PhD from the Technical University of Berlin. Contact him at dorgham.sisalem@tekelec.com.