SIP Vulnerabilities for SPIT, SPIT Identification Criteria, Anti-SPIT Mechanisms Evaluation Framework and Legal Issues

G.F. Marias, L. Mitrou, M. Theoharidou, J. Soupionis, S. Ehlert and D. Gritzalis

-------Pre-print version-------

19

SIP Vulnerabilities for SPIT, SPIT Identification Criteria, Anti-SPIT Mechanisms Evaluation Framework and Legal Issues

G. F. Marias
Athens University of Economics and Business

L. Mitrou
Aegean University

M. Theoharidou and J. Soupionis
Athens University of Economics and Business

S. Ehlert
Fraunhofer FOKUS Institute

D. Gritzalis
Athens University of Economics and Business

CONTENTS

19.1 Introduction 455
19.2 Background 457
19.3 SPIT Vulnerability Analysis 459
19.4 SPIT Identification Criteria 463
19.5 Anti-SPIT Mechanisms 466
19.6 Anti-SPIT Mechanisms Evaluations 470
19.7 Anti-SPIT Mechanisms and Legal Issues 474
19.8 Conclusions 475
References 476

19.1 Introduction

Since 2004 SPAM over Internet Telephony (SPIT) attack has been officially reported. The first incident was recorded in Japan at a major VoIP provider

called SoftbankBB, with Mio-subscribed customers*. Actually, SPAM history started 30 years ago, when the first e-mail SPAM was sent to 600 addresses. In the late 1970s and early 1980s there was not any real public Internet, and thus the growth of e-mail SPAM was not as exponential as it is today. Today, there is widespread use of the Internet and, since 2004, there has been significant growth in mass-market VoIP services over broadband Internet access services, in which subscribers make and receive calls as they would over the POTS (Plain Old Telephone Service) or PSTN (Public Switched Telephone Network). Thus, due to the current status of the Internet and the penetration of VoIP services, including the foreseeable IMS (IP Multimedia Subsystem) paradigm shift, it is likely we will see SPIT spread rapidly around the world. In fact, it is accepted to happen more sharply in the VoIP world than the e-mail world, since 'vishing' attacks, i.e. acquiring sensitive information, such as usernames, passwords and credit card details by masquerading as a trustworthy entity using VoIP services, are expected to increase by 50% during 2008, according to industry reports [1].

The explosion in the adoption and usage of VoIP together with the knowledge gained so far by spammers in the e-mail SPAM domain, and of course their potential profits of SPAM business, makes it more challenging to design and deploy an anti-SPIT framework. Additionally, VoIP services, and especially SIP (Session Initiation Protocol) when used as the signaling bearer for VoIP session management, significantly differ from classic SMTP e-mail services. Therefore, some of the existing and reliable mechanisms for prevention, detection and management of SPAM might not be applicable. Finally, even if we assume that SPAM and SPIT threats are equivalent, the SIP protocol itself might illustrate some new, probably more fatal, vulnerabilities and weak points. In this case, spitters might find themselves luckier than their predecessor spammers.

If SPIT prevalence becomes proportional to SPAM, then the acceptance of VoIP will be encumbered. This might be a problem from the service providers' point of view but the real problem will appear to the end-user, who will suffer simultaneously from spamming and spitting.

Even though SPIT is not yet a dominant Internet threat, several mechanisms and frameworks have been introduced to detect and counter the threat. Most of them combine or adapt existing and proved ideas from the anti-SPAM domain, while some introduce innovative new concepts. Until now as no practical amount of SPIT is present, their efficacy and efficiency could only be estimated in laboratory conditions, although a concrete evaluation framework is still missing.

Unsolicited communications infringe privacy, primarily in its narrow and visible sense, i.e. through the illegal intrusion into the private realms of the person. SPIT as a specific form of SPAM is considered to be an invasion of privacy [2]. VoIP SPAM (in the form of 'call SPAM', IM SPAM, Presence

*http://www.voipsa.org/pipermail/voipsec_voipsa.org/2006-March/001326.html. According to Columbia University officials (http://www.voipuser.org/forum_topic_10383.html) a recent SPIT accident was recorded also in the University during a VoIP pilot rollout.

SPAM), referred to as SPIT, differs from other 'traditional' forms of electronic communications (e.g. e-mail) in that is significantly more obtrusive and intrusive, as a phone will actually ring with every SPIT message, possibly even after midnight [3].

The fundamental right of privacy, anchored in various national constitutional texts as well as in the European Convention of Human Rights (Art. 8), encompasses informational privacy, relational privacy and freedom of communication in the meaning of privacy/secrecy of communications. Informational privacy (or right to informational self-determination) relates to the individual's right to decide autonomously, whether and which personal information they can communicate to others and/or processed by them. Affected is also the so-called relational aspect of privacy, i.e. the right to determine, which communications one wishes to receive or not [4].

Communication partners may reasonably expect that their communications and related data will not be used in an unlawful way by third parties. However, not only does SPIT constitute a threat to privacy, the protection against unsolicited communications through the use of various prevention and detection mechanisms raises a lot of questions concerning their compatibility with fundamental rights. Informational and communicational privacy and confidentiality are relevant with regard to preventing SPIT. Concerns are expressed in particular for existing practices to inspect communications in order to prevent and eliminate SPAM. Use of detection mechanisms and blocking of incoming calls/messages can restrict peoples' ability to communicate and therefore be an impairment of freedom of speech [4–6] and the right to receive and impart information, which is also recognized as an integral part of the freedom of information.

In this chapter we are discussing SPIT. We provide first a framework to defile and classify SPIT. Then, in Section 19.3, we identify some of the SIP vulnerabilities that might be exploited by potential spitters. We then stress the criteria that enable any anti-SPIT framework to identify which session and call-establishment attempts should be classified as SPIT. In Section 19.4 we present the results of our survey about the anti-SPIT frameworks already presented in the literature, as well as their classification as prevention, identification or handling approaches. Next we illustrate our proposed quantitative and qualitative criteria for assessment and evaluation of the proposed anti-SPIT frameworks, together with a compliance study. In Section 19.7 we discuss the legal issues pertaining to SPIT detection mechanisms focusing on protection of communications secrecy and privacy. Finally this chapter summarizes its findings and proposals in the corresponding conclusion section.

19.2 Background

19.2.1 SPIT definitions

SPAM over Internet Telephony is defined as a set of bulk, unsolicited calls or instant messages. A spitter uses the existing IP infrastructure to target users,

in order to initiate SIP calls or send SIP instant messages and generate SPAM. The phenomenon is already known from both the e-mail and the traditional telephony context.

SPIT can be identified in three different forms:

a) *Call SPIT*, which is defined as bulk, unsolicited session initiation attempts in order to establish a multimedia session.

b) *Instant Message SPIT*, which is defined as bulk, unsolicited instant messages and it is known as SPIM (SPAM over Instant Messages).

c) *Presence SPIT*, which is defined as bulk, unsolicited presence requests so that the malicious user becomes a member of the address book of another user or potentially of multiples users.

SPIT is a relative new identified threat; some first SPIT messages have been reported but the phenomenon has not yet reached the massive volume of the e-mail SPAM. The two types of SPAM (e-mail SPAM and SPIT) illustrate several common characteristics, like the use of the IP protocol to send e-mails and initiate calls or IMs, respectively. Attacks that are common in the e-mail context are applicable and expected to be seen in the SIP context too. Examples include harvesting addresses, dictionary attacks, zombies or bots.

19.2.2 Motivation

Despite the similarities that one can observe on the two SPAM types, there are also elements that differentiate the threats and vulnerabilities and justify the need for discovering new countermeasures to mitigate SPIT or adopt existing techniques cautiously. Firstly, SIP communication is not asynchronous as the e-mail one. An SIP call is held in real time and is sensitive to delays. Only at the call establishment phase some delay may be accepted and this renders the application of existing content analysis techniques to identify SPIT invalid. E-mail SPAM is mainly textual and may be combined by images, video or sound. SPIT takes the form of audio, video and maybe some instant text messages. The cost of sending a SPIT call is higher that sending an e-mail, in terms of resources and it can cause overload on a network. The annoyance of the user is considered higher when he receives a call as opposed to receiving an e-mail. Thus, one has to carefully examine the applicability of existing countermeasures to the SIP environment.

In order to design an anti-SPIT framework one needs to also identify the points where anti-SPIT countermeasures can be placed. These can firstly be applied on the domain of the spitter (i.e. outgoing proxy server) and act proactively in the sense that the SPIT message or call does not leave the boundaries of the caller's domain. In addition, mechanisms can be deployed in the domain of the user that is the target of SPIT (i.e. incoming proxy server), which act reactively and aim at identifying an incoming SPIT call or message. When SPIT is detected, handling mechanisms can be applied, which may vary from

flagging the call or message as potential SPIT to blocking it, depending on the domain's policies or the target's preferences.

To identify the overall amount of SPIT risk when using SIP for voice services, a vulnerability analysis is crucial. This study will report which of the SIP protocol specifications are vulnerable, whenever these specifications are mandatory or optional, if the interoperability with other protocols produces flaws, and, finally, if generic security weaknesses can be used for SPIT attacks. This is considered essential since potential anti-SPIT mechanism should be aware of these vulnerabilities in order to mitigate their impact.

Additionally, as several anti-SPIT mechanisms have recently been proposed in the literature, it is useful to survey their main functionality to identify metrics and criteria for their evaluation, in respect to SPIT. Actually assessment and evaluation of the existing anti-SPIT mechanisms is spread in many domains, such as efficiency (i.e. whenever it avoids SPIT false alarms or identifies the actual SPIT successfully), performance (i.e. overhead in the network, number of calls examined per minute, etc.), and finally, qualitative (i.e. scalability issues, adoption, availability, etc.). Another important dimension when evaluating existing or even forthcoming anti-SPIT mechanism is related to the legal framework and any anti-SPIT mechanism should be compliant to EC or national legislation.

19.3 SPIT Vulnerability Analysis

In the VoIP context, a spitter might perform various SPIT attacks by first building a set of active SIP URIs which contain callees' SIP address. Moreover, any call (i.e. INVITE) automation method can be useful for spitters, as well. As a conclusion, the identified SIP threats are separated into two main categories (a) creating a list of valid SIP addresses [1–7], and (b) when SPIT automation is required [8–10].

Sending ambiguous requests to proxies. A proxy server needs to decide the next step destination of a call or a message. The decision is made by looking at a specific part of the header of the request message or by contacting a service which is implemented by the SIP registrar server. If the header request/URI part does not offer adequate information then the proxy service is used. The proxy server offers an answer via a 485 SIP message. This kind of response message can contain a contact header field with a list of new URIs available for trial. For example if there are many URIs which contain a specific sequence of characters, then the 485 message would contain all the possible URIs with this string. The worst case arises where a special character can evoke a 485 message with all the registrars of the service's URI database. Therefore, this is an important weakness because a spitter can easily collect a set of valid URIs.

Listening to a multicast address. Every participant in an SIP session needs to be aware of the current location of other UAs. The physical location and the IP address of a user are discovered during the session initiation phase. In this phase, the proxy servers answers user or server REGISTER requests by consulting the location service. In order for the location service to have a valid URI database, it should be aware of any user change of address. This is achieved by listening to a specific multicast address where the UAs send a REGISTER message which contains their current IP addresses. The UAs listening to that address are therefore informed of users' location. On the other hand, if a spitter is listening to that multicast address, he is in the position to assemble both a set of the addresses of the registered users and their registration.

Population of 'active' addresses. Registration is a main function in the SIP infrastructure. The registration is accomplished when the REGISTER messages link the user's URI with the machine it currently uses and the IP address it happens to be using at that exact point in time. Therefore, both of the aforementioned weaknesses combined facilitate the determination of online users by spitters. This vulnerability allows spitters to send a smaller amount of SPIT messages and a larger number of successful SPIT since it can forward messages to users that are currently logged on.

Contacting a redirect server with ambiguous requests. As proxy servers can handle a great amount of messages; redirect servers are employed to balance the load of proxy servers. A redirect server is designed to refuse any other request than CANCEL. But it can collect a possible list of alternative locations of UAs and forward it to the requestor. The spitter can therefore send random requests to redirect servers and as a result collect a list of URIs in order to send SPIT messages. This attack could be more effective if conducted in conjunction with sending ambiguous requests to proxies (first vulnerability).

Throwaway SIP accounts. The creation of many accounts in various domains, by the same user, can prove to be a great vulnerability which can be exploited by spitters. This is because the SPIT traffic can not be easily associated with a specific SIP account, using conventional anti-SPIT countermeasures, such as Black Lists. This vulnerability is due to the SIP specification [7] which clearly states that a user can be registered in a SIP domain by simply sending a REGISTER message. Thus, a spitter is free to issue multiple REGISTER messages and as result to create multiple accounts in a SIP domain.

Misuse of stateless servers. Open Relay Servers, which exist in the email domain, are used by spammers to forward their SPAM emails. The SIP specification states that there can be stateless servers which offer the same service as Open Relay Servers in that they forward any call to its destination. Spitters can exploit it for hiding their IP address and location. Stateless proxies are not frequently used by SIP domains but are useful because they easily handle the rare flood of request messages.

Anonymous SIP service and Back-to-Back User Agents. The anonymous SIP option permits the forwarding of any call to a recipient without disclosing the caller identity. Spitters use this option to overcome various detection methods, such Black Lists and header content analysis. In this scenario a Back-To-Back User Agent (B2BUA) acts as a concatenated User Agent Server (UAS) and User Agent Client (UAC) [7]. The B2BUA collects and parses every request of a dialog [7]. Thus, a spitter can replace the contact information of every message with its own before forwarding the message. The recipient upon answering to his request would engage in a malicious dialog.

Sending messages to multicast addresses. The via header is one of the most important SIP header fields. The content of the via field shows the route that should be followed by the response message and the path the request has followed towards the target UA. The Maddr tag is a possible property of this specific field and indicates a multicast address. A spitter who obtains such an address is able to initiate calls to a set of users belonging to the multicast address.

Exploitation of forking proxies. A forking proxy server is a SIP proxy server which can send an INVITE message to multiple recipients simultaneously. These kinds of servers help spitters to collect a list of active SIP URI's faster, by using the vulnerability of sending ambiguous requests, which might result in a return of unambiguous new addresses. After a spitter sends an ambiguous INVITE request, it can initiate simultaneous calls towards all the addresses returned by the location service. In this case it is obvious that a spitter using one single request can automate bulk SIP calls generation utilizing a statefull proxy.

Exploitation of messages and header fields structure. A way of sending SPIT is through bulk unsolicited messages. Actually this could be achieved by (a) using the MESSAGE method, or (b) hiding the message in various SIP header fields of the SIP protocol message bodies. A spitter is capable of sending a SIP message, initiating a request, or spoofing the SIP responses transferred in the network. A request message might use the INVITE and the ACK methods, which both include a message body, which in turn might convey SPIT traffic. This body might include any media file that eventually will be delivered to the recipient. Furthermore, the MESSAGE method can be used to encapsulate SPIT traffic. This option might exploit the fact that any SIP flow outside a dialog does not require the authentication of the caller UA. On the other hand, the response messages include data fields that can be manipulated by a spitter. Table 19.1 illustrates the main response messages that can be used for SPIT content encapsulation:

Finally both request and response messages contain header fields that can conceal information which can be rendered by a UA. This kind of information can help spitters to send large messages since the attributes of these header fields allow (dynamic length, extent to multiple lines) or to perform other

TABLE 19.1
Response messages and their possible SPIT content.

Response messages	Possible SPIT content
Provisional (1xx)	The message bodies might include session descriptions
180 Ringing	The message body might contain textual information, or an audio file, or even animations
182 Queued	It might contain a reason phrase, illustrating further details about the status of a call, or a message body that play an audio file, from an on-hold music palette.
183 (Session Progress)	The Reason-Phrase or message body of it could provide additional details.
200 OK	It might contain returned information that depends on the method used in the request
300 Multiple Choices	It might include a message body
380 Alternative Service	It might contain a message body in which the alternative services are described.
480 Temporarily Unavailable	It might contain a reason phrase indicate a more precise cause, such as why the callee is unavailable
484 Address Incomplete	It might contain a reason phrase
488 Not Acceptable Here	It might contain a message body with a description of media capabilities
606 (Not Acceptable)	It might contain a message body

action as to trigger the UA to execute the body. The main header fields of spitters' interest and the way they can be exploited are:

***Subject*:** Might be parsed by the user's software and displayed to the user's terminal.

***From*:** Allows a text to be displayed to the user.

***Alert-Info*:** Specifies an alternative ring back tone that could be used as a pre-recorded audio SPIT message.

***Call-Info*:** Some parameters of the specific header, i.e., purpose, icon, info and card, could be used for sending SPIT messages.

***Contact and To*:** The display name flag could be altered to provide a SPIT message.

***Retry After*:** The optional textual comment parameter might be used to convey a SPIT message.

***Error-Info*:** The pointer to additional information, in relation to the error status response, could be set by an attacker to point to a SPIT message.

***Warning*:** The warning text that this header contains could be replaced by a SPIT message.

***Content-Disposition*:** Even if this header field indicates the way that a message should be interpreted by a UA, it could also be used as a SPIT message.

***Content-Type*:** This field denotes the media type of the message body. Therefore, it could be used as a SPIT message. It could also trigger the receiver's UA to execute the message body.

***Priority*:** This field indicates the urgency of a request. A spitter could exploit it in order to facilitate the delivery of the SPIT message.

There are two more vulnerabilities; one due to the protocol's optional recommendations and the other due to the interoperability with other protocols. The first vulnerability is owed to the exploitation of registrar servers. The registrar server is in charge for updating the location information of the UAs by handling REGISTER requests. A server answers to queries of UAs and discloses the location of registered users. Therefore, a spitter can issue an attack to registrar in order to collect a set of URIs. This attack is successfully accomplished because the user, who initiates a dialog to a registrar, is not compulsory registered or authenticated to a domain since the SIP specification does not oblige the domain to register each user. Thus, the impersonation of a legitimate user is easily acquired. The second vulnerability is the exploitation of particular domains' address resolution procedures. The SIP proxy servers and the afterward communication inherits the vulnerabilities which are introduced by the communication supplementary protocols, such as DNS queries, ARP, RARP, etc. These protocols can be vulnerable to several attacks, such as spoofing or man in the middle [8–10]. Therefore, a malicious intermediate could impersonate a legitimate user, spoof messages, deliver SPITs by using the victims' identities and create lists of other possible victims.

19.4 SPIT Identification Criteria

A list of SPIT identification criteria are analyzed in this section. These may be applied as SPIT detection rules on both sides of a SIP communication and they can be proactive or reactive depending on the point they are applied (spitter's or target's domain). We propose two generic categories of SPIT identification criteria:

- ***SIP Message criteria:*** This category includes the criteria that are related to attributes of SIP messages and they can be based on (a) call and message patterns or (b) headers' semantics.

- ***SIP User Agent criteria:*** This category includes the criteria that are related to attributes of a SIP User Agent and they can examine the origin of the call or message as well as the relationship of the SIP participants.

19.4.1 SIP Message Criteria

19.4.1.1 Call-Messages Patterns

This category includes criteria that analyze specific call or message characteristics or patterns, in order to determine whether a call (message) is possibly a SPIT call.

- *Path traversal:* A call or a message might pass through many intermediates before reaching its final destination. This path is denoted in the *via* header. Thus, if in the via header a SPIT domain is recognized, the call or the message may be a potential SPIT.
- *Number of calls-messages sent in a specific time frame:* It analyzes the number of calls (messages) made in a specific time period by a user. If this number is above a specific pre-defined threshold then the call (message) is characterized as a possible SPIT call.
- *Static calls' duration:* If the calls initiated by a single user have a static duration, then the user is a potential spitter who possible used an automated script in order to initiate the calls.
- *Receivers' address patterns:* If the receivers' addresses follow a specific pattern (e.g. alphabetical SIP URI addresses), then the call (message) is flagged as potential SPIT.
- *Small percentage of answered/dialed calls:* It indicates the number of successful call completions from this caller per a pre-defined time period, which is relative to the number of failed ones.
- *Large number of errors:* When a user sends a large number of INVITES and the SIP protocol returns a large number of error messages (e.g. 404 Not Found) is a sign of a potential SPIT attack.
- *Size of SIP messages:* In this case a set of SIP messages sent by a user to other users is analyzed. If those messages have a specific size then it is very possible to be sent by a 'bot' software, and therefore the call is characterized as SPIT.

19.4.1.2 SIP Headers' Semantics (SIP Message Oriented)

This category includes criteria that identify a SPIT call or message through a semantic analysis of the contents of the SIP messages (e.g. Bayesian filtering). These particular criteria are further categorized, according to the different parts of SIP messages that could be used. These are: (a) a message's headers,

(b) a message's body, and (c) the reason phrases of a message. In addition, we have identified three possible types of SPIT that could be injected in a SIP message, namely: (a) text SPIT injected in a header field, (b) media SPIT carried in the message body, and (c) hyperlink SPIT injected in a header field.

Tables 19.2 through 19.4 depict the specific SIP header fields that can be used for a detailed semantic analysis, so as to detect a SPIT call or message alongside with the type of the SPIT that could be sent.

Table 19.3 presents the type of messages containing a body, which could be used in order to make a SPIT call (or message). They are categorised in terms of request and response messages. SPIT contained in the message body can be text, media or hyperlink.

TABLE 19.2
SIP headers that could be used for SPIT.

Header fields	SPIT type	Request messages	Response messages
Subject	Text	✓	✓
From	Text	✓	✓
Call-Info	Hyperlink	✓	✓
Contact	Text	✓	✓
To	Text	✓	✓
Retry After	Text	✓	✓
Alert-Info	Hyperlink	✓	✓
REPLY TO	Text	✓	—
Error-Info	Hyperlink	—	✓
Warning	Text	—	✓
Header fields related to SIP messages' bodies (not carrying SPIT 'directly')			
Content-Disposition	Displayed message body	✓	✓
Content-Type	Displayed message body	✓	✓

TABLE 19.3
Request–response messages that could be used for SPIT.

	INVITE
Request messages	ACK
	180 Ringing
	183 Session Progress
	200 OK
Response messages	300 Multiple Choices
	380 Alternative Service
	488 Not Acceptable Here
	606 Not Acceptable

TABLE 19.4
Reason phrases that could be used for SPIT.

182 Queued
183 Session Progress
200 OK
400 Bad Request
480 Temporarily Unavailable
484 Address Incomplete

Table 19.4 presents the reason phrases of response messages that could be used for SPIT purposes. Reason phrases may contain text or hyperlink type of SPIT.

19.4.2 SIP User Agent Criteria

These criteria examine the characteristics of a SIP session, meaning the SIP addresses of the sender/caller (i.e. SIP URI or IP address), as well as the domain the session was initiated in. They can be used in traditional white and black listing techniques.

- *Caller SIP URI*: It detects and analyzes the SIP URI of the sender of a call/message, so as to determine if he/she is a potential spitter or not.

- *Caller IP address:* It analyzes the IP address of the sender/caller so as to characterize him/her as a spitter.

- *Caller domain:* It analyzes the identity of the domain of the caller (sender), which is determined either by SIP URI of the caller, or through DNS lookup from the IP address. If the identity of the domain is a well-known SPIT source, then the call or the message is characterized as potentially SPIT.

- *Caller/callee relationship:* It examines whether the caller/sender is trusted by the callee/receiver. Typical examples include whether a caller is known to the callee (inclusion in address book, previous calls have been established, marked as spitter by the callee).

19.5 Anti-SPIT Mechanisms

19.5.1 Anti-SPIT Mechanisms Description

SPIT is likely to influence the future use and adoption of the VoIP technology. In order for SPIT to be efficiently managed the following three progressive steps are necessary: (a) prevention, the avoiding of SPIT altogether,

(b) detection, the ability to identify a SPIT call or message, and (c) handling, the dealing with a detected SPIT call or message.

Certain general frameworks from the e-mail SPAM paradigm have been considered as potential candidates for SPIT avoidance [8]. Some of them appear to be basic building blocks of the anti-SPIT architecture that have been proposed in the literature. We shall first provide a brief description of the general anti-SPIT frameworks before proceeding to classify them according to the three aforementioned progressive steps.

Black and white lists. White lists are made up of trusted users that are not categorised as spitters. An end-user accepts the calls, or messages, initiated by any of the members in his/her white list. On the contrary, black lists contain any/all potential initiators of SPIT calls. These calls are therefore to be blocked.

Content filtering: This method is based on filters that scan the content of messages. They appear to be inappropriate as anti-SPIT, since real-time filtering is hard to accomplish. Nevertheless, this technique could be used for the detection of instance messaging SPIT (similar to e-mail SPAM).

Challenge–response. Communication becomes possible only once the caller has replied correctly to a challenge sent by the callee. This approach aims at preventing SPIT by operating a distinction between humans and 'bots'. Other such mechanisms include Turing tests and computational puzzles.

Consent-based. Here, communication is not achieved unless the callee explicitly consents.

Reputation-based. Trust is the central notion of this approach. When a callee receives a request for communication the level of trust of the caller should be determined. This is accomplished through direct estimations or second-hand reputations. If the trust level is above a predefined threshold then the communication is permitted, otherwise it is rejected.

SIP addresses management: One of the spitters' main goals is to collect as many valid addresses as possible. So, to prevent SPIT it is important for end-users to protect their addresses (i.e. URIs) from being collected. For this to be done, two different approaches are possible: address obfuscation and use of multiple addresses.

Charging-based. This approach obliges spitters to pay for every unsolicited bulk call (messages), as a result, the cost born by the spitter is increased in terms of financial or computational resources.

As more attention is drawn to the SPIT problem, a number of anti-SPIT frameworks begin to emerge. They are commented upon briefly in the sequel.

19.5.1.1 SPIT Prevention using Anonymous Verifying Authorities (AVA) [11]

The so-called AVA system illustrates how the prevention of voice SPAM may be achieved by extending the call setup procedure. This method is founded

upon a 'call me back scheme' and the use of two new entities in the IP infrastructure, namely: (a) the mediator, and (b) the AVA. Through the exchange of information between these two entities, SPIT is mitigated by anonymously blocking unwanted calls. This way the caller is not aware of the existence of the callee, provided that the call failed to be established.

19.5.1.2 SPIT Mitigation through a Network Layer Anti-SPIT Entity [12]

The anti-SPIT entity is based on an approach which detects and mitigates SPIT by using a sniffing-oriented network-level entity. This entity filters and analyses the network traffic so as to detect a SPIT call by means of educated guesses. These guesses are based on a list of criteria, which contribute to different extent toward the final decision. The mechanism only takes the SIP packets into account and ignores the rest of the network traffic. Finally, handling actions are enforced based on the results of the SPIT analysis. These actions rely on the policies adopted by the domain to which the specific user belongs, and the end-user's preferences.

19.5.1.3 SPIT Detection based on Reputation and Charging Techniques [13]

This approach proposes a SPIT detection mechanism based on two different techniques, namely reputation (i.e. trust networks) and payments at risk. The reputation-based technique is based on the notion of trust, in other words, SPIT detection relies solely on the callee's trust toward the caller. On the other hand, the charging based technique aims at reducing SPIT by increasing its cost. This is achieved by imposing a charge to each message sent by the caller. The specific technique functions alongside simultaneously with other anti-SPIT frameworks such as authentication modules and white lists.

19.5.1.4 DAPES [14]

The DAPES system determines in real time depending on whether a call is identified as being SPIT or not. The main characteristics of DAPES are the following: firstly, all messages are sent through proxies that serve as authenticators, secondly, all outbound proxies have certificates granted by Trusted CA, thirdly, all communication between proxies has to be encrypted and fourthly, the sender has to be authenticated by its domain proxy. In order for all of the above to be effective, certain characteristics have to be supported. Current systems are therefore unable to effectively implement the abovementioned infrastructure.

19.5.1.5 Progressive Multi Grey-Leveling [15]

Progressive multi grey-leveling (PMG) determines and allocates a grey-level for each caller that enables it to determine whether a call qualifies as SPIT or

not. This mechanism is not based on the feedback of other users with regards to a specific caller but on the last calls made by that caller itself. In practice, the grey level of each user is calculated by means of the addition of a long-term and short-term level. If a summary is greater than a threshold than the user is considered as being a spitter and all his calls are consequently blocked. However, a caller who is classified as a spitter is bound to loose such status since the grey-level of the caller is not constant.

19.5.1.6 Biometric Framework for SPIT Prevention [16]

This technique offers an approach based on identity-management since a spitter changes his identity frequently. The specific technique proposes that personal detail (such as biometric data) of all users be recorded, so as to create a unique link between every user's identity and its biometric data. So as to enable the functioning of this infrastructure, each user must be registered to authenticated servers when using VoIP. This technique requires an effective communication between servers that can enable the exchange of user credentials (based on biometric data).

19.5.1.7 RFC 4474 [17]

This proposal focuses on identity-management and aims at resolving the problem of user authentication by using PKI and certificate authorities. So as to enable the functioning of this RFC, two new sip-header fields are required: (a) identity, containing a signature that is used to verify the caller's identity (b) identity-info, for transmitting a reference to the signer's certificate. Despite the fact that this infrastructure had not been initially proposed for SPIT, it can nevertheless be easily implemented to SPIT.

19.5.1.8 SIP SAML [18]

The Security Assertion Markup Language (SAML) is used for the expression of security assertions, such as authentication, role membership, or permissions. A SIP-authentication service is proposed that authenticates the user via certain asserted features. Each VoIP message includes caller identity information along with a reference to a SAML assertion which has various features of the caller and the caller's domain certificate. This infrastructure tries to avoid the frequent change of spitter addresses through identity control.

19.5.1.9 DSIP [19]

Differentiated SIP is an extension to the SIP protocol which classifies users into three distinct categories. These categories are deduced from the e-mail context: the white list, made up legitimate callers, the black list comprises the spitters and the grey-list contains the callers who have yet to be categorised. DSIP usually employs human verification test for the uncategorised callers.

Once the grey-list callers have succeeded the test, the communication with the callee is established.

19.5.1.10 Voice SPAM Detector [20]

The voice SPAM detector system is a combination of anti-SPIT techniques that is founded upon reputation and trust. The basic characteristics of a VSD system are the following: (a) Presence filtering that depends on the current status of the callee, (b) the traffic patent filter evaluates the incoming calls received from a specific caller or ship domain so as to ensure that they do not exceed the predetermined threshold, (c) the black and white list that is based on the black and white list filtering, (d) Bayesian learning, every call is evaluated, on the basis of trusted information that is made available by third party entities, with regards to the behaviour of the participating entities. The existence of trusted information is based on the presumption that all participating entities have established prior calls, (e) social networks and reputation: this technique is used so as to accept a call that is based on social relationships already established by the user in its VoIP environment.

19.5.1.11 VoIP SEAL [21]

VoIP SEAL is a system that is divided into two main stages. The first stage uses various models which give a score between –1 and 1, the higher the score, the higher the probability of the call being SPIT. Moreover, two thresholds exist: On the one hand, if the score is inferior to the lower threshold, the call moves to a second stage that usually entails a CAPTCHA test. On the other hand, if the score is higher than the superior threshold the call is rejected. The main characteristic of VoIP SEAL is the adoption of a modular architecture that facilitates adding or updating of certain modules so as to take defence measures against SPIT.

19.5.2 Anti-SPIT Mechanisms Classification

AU: Check Meaning/Intent

Classification of anti-SPIT countermeasures in prevention, detection and handling is worthy, because it provides information about the scope. For example, it might be useful to combine prevention with handling mechanism, or detection with handling mechanism.

In the following prevention refers to the mechanisms applied to avert SPIT, i.e. avoid a SPIT call reaching the callee domain or phone. This avoids communication and processing overheads, conserves costs and resources, and minimizes the annoyance of the end-user. Detection means the identification of SPIT when a SPIT call is actually under transit to the callee, or under processing to the callee's proxy server or SIP-phone. The issue here is to mitigate or avoid any annoyance, and if possible, to minimize the overheads and to conserve resources. Finally, handling means any reaction to the SPIT. Table 19.5 illustrates this classification.

TABLE 19.5
Classification of the anti-SPIT mechanisms.

Mechanism	Prevent	Detect	Handle
AVA [5]	√		
Anti-SPIT Entity [6]		√	√
Reputation Charging [7]	√	√	
DAPES [8]	√		
PGM [9]		√	
Biometrics [10]	√	√	
RFC 4474 [11]	√		
SIP SAML [12]	√	√	
DSIP [13]	√	√	
VoIP SEAL [14]	√	√	√
VSD [15]	√	√	

19.6 Anti-SPIT Mechanisms Evaluations

19.6.1 Assessment Criteria

An assessment of the proposed anti-SPIT mechanisms requires the definition of various qualitative and quantitative criteria. We have based our research on the following criteria:

Percentage of SPIT calls avoided. How many SPIT call attempts have been identified and handled by the anti-SPIT mechanism?

Reliability. The precision of making the right adjustments about SPIT calls and callers, in terms of false positive and negative rates.

Latency. Due to the real-time nature of VoIP, quick decisions regarding SPIT detection are a major requirement, especially when legitimate calls are analyzed. In such a case, trustworthy users should not tolerate large delays.

Human interference. This metric represents the transparency of the anti-SPIT mechanisms to the end-user.

Resource overhead for the SIP provider. SIP providers should estimate the required resources for the implementation of the mechanism. This quantitative criterion seems essential for providers, since the number of calls that should be analyzed per unit time might be enormous, when the number of registered users increases.

Vulnerabilities. This parameter refers to the capability of a spitter to bypass any of the anti-SPIT countermeasures.

Privacy risk. This criterion is associated with the collection, manipulation, and dissemination of private data. We assume that the end-user consents for

the collection and manipulation of his/her private data, and has authorized specific legal entities for these purposes.

Scalability. This is an important criterion, since VoIP networks grow fast. Scalability should be considered when authentication is involved, since PKI and CA might need to establish complex cross-certification chains, or when reputations and assertions are used.

Adoption. This parameter corresponds to the success of the anti-SPIT countermeasure, and depends on the effort it takes for an end-user, or a provider, to begin using it.

Availability. It denotes the increase in the availability of network, computing, memory, or human resources when preventing, detecting or handling SPIT countermeasures apply.

More specifically, in the description of VoIP SEAL mechanism [21], it is mentioned that the particular mechanism is intrusive for the end-user, as there is an interactive part requiring user's feedback. In our analysis, we do not estimate how intrusive this mechanism is or not, but we emphasize that there are information in the description of the mechanism that could help someone value accordingly the particular criterion, namely human interference. Finally, regarding the vulnerabilities criterion we must emphasize that we only focus on the vulnerabilities considerations that the authors take into account, whether these are few or many.

19.6.2 Compliance of SPIT Mechanisms to Assessment Criteria

Table 19.6 presents which mechanisms takes into account the SPIT identification criteria we defined. For this purpose we only took into consideration an abstract description of each mechanism. Furthermore, we do not consider whether the mechanisms meet the criteria well or not, but we rather provide the mere existence of each criterion in the mechanisms' description. For example, in the description of reputation/charging mechanism, the use of black and white lists requires the existence of a way to identify and handle users, either by SIP URI, IP address or even domain of origin. However, as something like that is not explicitly mentioned, we put the appropriate negative value in the table. Furthermore, the table can be used as a reference to choose the appropriate mechanism for anti-SPIT mechanisms in a given context. For example the call and message patterns might be highly cost demanding, in terms of data gathering and analysis, and thus mechanisms that focus on, and better fulfill the other criteria might be of preference. Finally, the table can be read as a concentrated area of further research directions regarding anti-SPIT countermeasures. Some of the questions that one can answer using the table include how can a particular mechanism contribute in terms of prevention, detection or handling of SPIT, and which combinations of techniques should someone use in order to fight SPIT more effectively, etc.

TABLE 19.6
Compliance of SPIT mechanisms to identification criteria.

	SPIT identification criteria										
	Call and message origin			Calls/messages patterns							Caller/ callee relationship
Anti-SPIT mechanisms	Caller SIP URI	Caller IP address	Caller domain	Path of message	Number of calls	Calls duration	Sequential call numbers	Dialed/ answered calls	Large number of errors	Size of SIP message	White/ black lists
AVA	√	–	–	–	–	–	–	–	–	–	–
Anti-Spit Entity	√	√	–	–	√	√	√	–	√	–	√
Reputation/ Charging	–	–	–	–	–	–	–	–	–	–	√
DAPES	√	√	√	√	–	–	–	–	–	–	–
PMG	√	–	–	–	√	–	–	–	–	–	√
Biometrics	–	–	–	–	–	–	–	–	–	–	–
RFC 4474	√	–	√	–	–	–	–	–	–	–	–
SIP SAML	√	–	√	–	–	–	–	–	–	–	–
DSIP	√	–	√	–	–	–	–	–	–	–	√
VoIP Seal	–	–	–	–	–	–	–	–	–	–	√
VSD	√	√	√	√	√	–	–	–	–	–	√

Finally it must be mentioned, that although many mechanisms check the content of the SIP messages, for example the headers *FROM, VIA*, etc, none of them check the appropriate the messages' bodies and reason phrases for carrying the actual SPIT message. This explains why the corresponding column in the table has only negative values.

19.7 Anti-SPIT Mechanisms and Legal Issues

Protection of individual rights has—or should have—an impact on the choice and design, implementation as well as on the legal assessment of anti-SPIT measures and mechanisms. The deployment of anti-SPIT mechanisms raises a lot of issues relating to everyone's right to respect to his private life and his correspondence. The confidentiality of communications is guaranteed explicitly by several national, supranational and international legal instruments [among them: the Fourth Amendment (USA), the EU e-Privacy Directive (Art. 5), the European Convention for the Protection of Human Rights and Fundamental Freedoms (Art. 8)]. Both the EU e-Privacy Directive and the European Convention for Human Rights prohibit any form of interception, e.g. a third party acquiring access to the content or traffic data (data processed for the conveyance of a communication or the billing thereof) to private communications between two or more correspondents. Such interceptions are acceptable only on the basis of some fundamental criteria, i.e. legal basis, need for such a measure in a democratic society and conformity of the measures adopted with legitimate aims such as national security, public safety or the economic well-being of the country, prevention of disorder or crime, protection of health or morals, protection of the rights and freedoms of others.

Electronic communications are combining, in the most regular cases, both the notions of 'private life' and 'correspondence' [5]. Interception, opening, reading, delaying reception of communications or impeding the sending of messages have been considered by the Courts to be an intrusion into the right of correspondence, which includes not only confidentiality but also the right to send and receive correspondence. In this institutional context anti-SPIT mechanisms have to respond to a double challenge: they have to prevent successfully the invasion into the receiver's privacy while respecting the privacy and other fundamental rights of communicating parties, even of potential spitters.

Filtering and withholding of received communication constitutes an interference with the freedom of communications. Filtering and screening of communications content for the purpose of detecting SPIT, without the consent of the communicating parties, should be considered as intruding fundamental rights and therefore unlawful. In any case defining SPIT by content is related with the risk of infringing freedom of expression and introducing a kind of censorship. Especially some techniques of communication screening,

like blacklisting, can raise questions in relation to the freedom of expression and freedom of information [5]. Filtering can also result in the blocking of legitimate information (the so-called false positive), which has as consequence the infringement of the freedom of expression [2,5].

Blocking SPAM by technical means could intrude on the right to informational self-determination of the communicating parties. By definition some anti-SPIT modules need information from the user, such as his preferences, to be able to work [22]. Mechanisms based on the analysis of behaviour and reactions (such as the challenge–response mechanism), the caller's characteristics reputation or the use of black/white lists imply the collection and use of personal data, i.e. any information relating to an identified or identifiable individual. Phone numbers allow indirect identification of subscribers through the use of reverse directories as well as through electronic communications services providers. Indirect identification is also possible through IP addresses, which can be traced back to a computer and through the provider consequently to a subscriber. Even if the link between subscribers and users is, in the case of IP address, less strong than by e-mail addresses and phone numbers, most IP addresses can be tied to a log-in and may qualify as personal data [2].

This challenge–response mechanism can have as result the processing of personal data of spitters to the extent that these data can be considered as personal. The authentication and reputation manager mechanisms are based on the calling party reputation as well as on the enhancing identity management of the SIP users and consequently they may imply the use of personal data. In the case of black/white lists, the SIP URL of the users is stored and consequently it could be used for indirect identification. The list of 'friends' and 'non-friends' are also stored by their URIs, revealing a user's personality and life profile. In the aforementioned cases the processing of personal data should be based either on the explicit consent of the user, as far as his data are processed, or on the protection his right to protect his privacy, which regularly overrides the interests of the spitter. Filtering tools may not be in compliance with the existing data protection legislation. Subscribers should keep the control over the information concerning them by having the possibility to opt out of SPIT detecting and the possibility to decide, what kind of SPAM should be filtered out.

19.8 Conclusions

The SIP protocol raised significant concerns, as to whether SPIT will be equivalent to the current SPAM prevalence. In order to address and evaluate these concerns, we addressed existing vulnerabilities of SIP for SPIT and proposed a SPIT identification framework. Additionally, we provided a macroscopic view of SPIT management techniques and mechanisms, alongside

an extensive list of evaluation criteria that can be used for self-assessment against SPIT. Moreover, we presented a high-level and theoretical evaluation of the existing anti-SPIT mechanisms, and, finally we addressed the legal issues that are raised by the deployment of these candidate anti-SPIT mechanisms.

References

[1] McAfee Avert Labs. 2007. Top 10 threat predictions for 2008.

[2] Asscher, L.V., and J. van Erve (IViR -Institute for Information Law). 2004. Regulating spam–directive 2002/58 and beyond. Amsterdam, 1–74.

[3] Posegga, J., and J. Seedorf. 2005. Voice over IP: Unsafe at any bandwidth?

[4] Kabel, J. 2003. Spam: a terminal threat to ISPs?. *Computer und Rrecht International* 1:1–5.

[5] Data Protection Working Party, Opinion 2/2006 on privacy issues related to the provision of e-mail screening services (WP 118).

[6] Sury, U. Datenschutzgerechte nutzung von intrusion detection–Organisatorische und juristische implikationen. DuD 29 (2005) 7: 393–98.

[7] Rosenberg, J. 2002. SIP: Session Initiation Protocol. RFC 3261 June.

[8] Rosenberg, J., and C. Jennings 2006. The Session Initiation Protocol (SIP) and Spam. Draft-IEFT-sipping-spam-03 October.

[9] Kuhn, D., T. Walsh, and S. Fries. 2005. Security considerations for voice over IP systems. Special Publication No. 800-58, NIST, USA.

[10] Sisalem, D., J. Kuthan, and S. Ehlert. 2006. Denial of service attacks targeting a SIP VoIP infrastructure: Attack scenarios and prevention mechanisms. *IEEE Network Journal* 20(5): 26–31. September–October.

[11] Croft, N., and M. Olivier. 2005. A model for spam prevention in voice over IP networks using anonymous Verifying Authorities. In *Proceedings of the 5th annual information security South Africa conference (ISSA 2005)*. South Africa. July.

[12] Dantu, R., and P. Kolan. 2005. Detecting spam in VoIP networks. In *Proceedings of steps to reducing unwanted traffic on the Internet workshop (SRUTI '05)*. USA. July.

[13] Dritsas, S., J. Mallios, M. Theoharidou, G. F. Marias, and D. Gritzalis. 2007. Threat analysis of the session initiation protocol regarding spam. In *Proceedings of the 3rd IEEE international workshop on information assurance (WIA 2007)*. USA. April.

[14] Srivastava, K., and H. Schulzrinne. 2004. Preventing spam for SIP-based instant messages and sessions. Technical Report, University of Columbia.

[15] Dongwook, S., A. Jinyoung, and S. Choon. 2006. Progressive multi gray-leveling: A voice spam protection algorithm. *IEEE Network Journal* 20(5): 18–24. September–October.

[16] Baumann, R., S. Cavin, and S. Schmid. 2006. Voice over IP–Security and SPIT. Swiss Army, FU Br 41, KryptDet Report, University of Berne, September.

[17] Peterson, J., and C. Jennings 2006. Enhancements for authenticated identity management in the session initiation protocol. RFC 4474, August.

[18] Tschofenig, H., R Falk, and J. Peterson. 2006. Using SAML to protect the session initiation protocol. *IEEE Network Journal* 20(5): 14–17. September–October.

[19] Madhosingh. 2006. The design of a differentiated SIP to control VoIP spam. Technical Report, Computer Science Department, Florida State University.

[20] Niccolini, S. SPIT prevention: State of the art and research challenges. NEC Europe, Germany: Network Laboratories.

[21] Dantu, R., and P. Kolan. Detecting spam in VoIP networks. In *Proceedings of steps to reducing unwanted traffic on the Internet workshop (SRUTI '05)*. USA. July 2005.

[22] Rohwer, T., C. Tolkmit, M. Hansen, M. Hansen, J. Moeller, and H. Waack. Abwehr von spam over internet telephony. (SPIT-AL), Kiel 2006. 1–30.