Lightweight Security Mechanism for PSTN-IP Telephony Convergence

Wojciech Mazurczyk [†] **and Zbigniew Kotulski**[††]

[†]Warsaw University of Technology, Institute of Telecommunications, Poland
[††]Institute of Fundamental Technological Research of the PAS, Warsaw, Poland

Summary
A new, lightweight security mechanism for the PSTN-VoIP hybrid networks is described. It is based on the two information-hiding techniques: audio watermarking and network steganography. The proposed scheme is suitable especially for the PSTN-IP-PSTN (toll-by-passing) scenario, which nowadays is a very popular application of the IP Telephony systems. Proposed mechanism can be used to authenticate end-to-end transmitted voice between PSTN users. Additionally it improves IP part traffic security (both media stream and VoIP signalling messages). Exemplary scenario is presented for SIP signalling protocol along with SIP-T extensions and H.248/Megaco protocol.

Key words:
VoIP security, PSTN-IP Telephony interconnections, network steganography, audio watermarking

1. Introduction

Currently, PSTN (Public Switched Telephone Network) voice services are characterized by three important parameters: good voice quality, high reliability and security. We are able to provide critical services, such as emergency numbers or federal agencies with ability for the lawful intercept. All of these issues should be addressed before VoIP (Voice over Internet Protocol) systems are deployed on a mass scale and security become the most critical area [5]. With the deployment of the IP Telephony accelerating, comes the increasing need to find ways to effectively secure such systems and to make it as secure as PSTN.

Nowadays IP networks security solutions are usually based on the deployment of a number of the security devices and applications to protect and monitor networks, such as: firewalls, Intrusion Detection (Prevention) Systems (IDS/IPS), Virtual Private Networks (VPN), authentication services, anti-virus software, etc. Paradoxically, VoIP is highly sensitive to the delay, packet loss and jitter, making these security mechanisms inadequate [5]. That is why in the IP network environment, especially for the real-time services like VoIP, it has become desirable to develop security mechanisms that efficiently utilize available resources.

Another challenge is that there must be cooperation between security measures for the IP Telephony for different signalling protocols that VoIP can be based on. The lack of this cooperation makes protecting of the IP Telephony service ineffective from more sophisticated attacks or internal threats. Let us consider a scenario where calls are sent through the PSTN-IP-PSTN network (which is nowadays one of the most popular application for the IP Telephony system). In this situation we must point out that the IP part is the weaker one from the security point of view. In this case scenario by the term "IP part" (or IP backbone) we do not necessarily mean an original (private) service provider's IP network. An example of the Skype [9] shows that one can become serious telecom provider almost without his own infrastructure but with the use of the Internet resources. In this paper we assume that the IP part can be a potential threat for the PSTN-IP-PSTN; that is why it needs to be secured.

The convergence between those two types of the telephony networks is unstoppable. Nowadays a vast number of the telecom providers offer telephony connections from and to PSTN/VoIP networks. In this scenario new security constrains emerge. That is why we propose a new, lightweight security mechanism for PSTN-VoIP. It is based on audio watermarking and network steganography and it establishes a reasonable trade-off between security and performance even for a low-bandwidth environments.

2. Motivation

We would like to emphasize that it is not enough to protect only the media streams exchanged between the calling parties. If we limit IP telephony security only to this case we may witness a situation, in which the two entities are able to talk to each other in a secure manner, but they are unable to initiate a call because of an attacker's actions on the signalling protocol. Usually, the users do not realize how much the weak security of the signalling protocol, or its absence, can impact the security of the whole system. They only often demand their

Manuscript received May 5, 2007
Manuscript revised May 20, 2007

conversations to be confidential and not revealed to a third-party. Thus, both types of the VoIP traffic: signalling messages and media streams need to be secured [5].

As mentioned in Section 1, the real problem for the convergence between legacy and new voice systems is securing the IP Telephony part of the PSTN-IP-PSTN connection. We should "seal" VoIP security gaps, so it would not affect the PSTN part. That is why, in this paper, we are proposing a different approach to the PSTN-VoIP security, based on the audio watermarking and network steganography. Our motivation to increase security for the PSTN-VoIP interconnections is based on the following facts:

- There is a need for security mechanism for the PSTN-IP-PSTN scenario to adjust security of the IP part to the PSTN security level,
- There is a need for a low-power computing and low-bandwidth consuming mechanism for the real-time services, like VoIP, to ensure a certain level of security without affecting performance. As showed in [10], security solutions like VPNs are not a suitable for securing VoIP traffic - the price of the strong security is a decisive drop in the QoS parameters. Also there are drawbacks of the SRTP protocol which are discussed in [4]; SRTP is the most popular mechanism to provide authentication and integrity for the data stream,
- There are no standardized security solutions for the interworking between different signalling protocols, e.g., ISUP and SIP (PSTN-VoIP), or even SIP and H.323 (VoIP-VoIP). Every signalling protocol uses a disjoint set of the security mechanisms. In the PSTN-IP-PSTN scenario we must ensure that the PSTN signalling messages, as well as conversation, will be transmitted in a secure manner during travelling through the IP environment,
- It is not easy task to secure VoIP system based on one VoIP signalling protocol [5] and providing secure interworking in PSTN-VoIP scenario is much more complex,
- VoIP security is still evolving. Lately, a large number of the new security mechanisms were standardized for IP Telephony, especially those for securing signalling messages, but they often have certain disadvantages (like the big overhead in S/MIME for SIP or the use of the TLS only for the TCP, whereas traffic is carried mostly by UDP [5, 11]). So, the process of the developing new mechanisms is not finished and still it is time for new solutions and ideas, especially for securing PSTN-VoIP interworking.

The rest of the paper is organized as follows. In Section 3 a general idea of the proposed solution is presented. Then, in Section 4 and 5, the mechanism's operations in IP network and PSTN part of the PSTN-IP-PSTN scenario are presented. Next, in Section 6 security services that we gain with the use of proposed solution are outlined. Finally, we end with conclusions in Section 7.

3. General idea of the proposed solution

We proposed using digital watermarking and network steganography techniques to improve VoIP security in [1], [2] and [3]. In [1] we used digital watermarking to authenticate VoIP conversation as well as messages of the signalling protocol that IP Telephony is based on. In [2] we used a covert channel created in voice to alternate RTCP protocol functionality and additionally to provide security of the VoIP conversations. In this paper we combined the two information-hiding techniques mentioned earlier. In [3] a general description of the protocol that uses multipurpose covert channel in VoIP transmission was defined. The network steganography was used to transmit the header and the digital watermarking was used to transmit payload of the protocol's PDU (Protocol Data Unit).

Here we propose to combine proposed solutions to improve security in the PSTN-IP-PSTN scenario. The main idea is to authenticate conversation end-to-end (between PSTN users) with the watermark embedded into the voice and additionally use network steganography and digital watermarking in the IP environment to ensure VoIP signalling protocol and ISUP (PSTN signalling protocol) security. Because we assume that the IP network (of the PSTN-IP-PSTN) can be public (e.g. Internet) that is why it is vital to make sure that we can secure IP part efficiently and adequate.

So, in this solution we will embed two watermarks into the voice stream: one at the PSTN endpoint and another at the entrance to the IP part of the network. The watermarking technique that is used must allow to embed/extract one of the watermarks without distortion or destroying of the other existing ones. That means that we will be watermarking already watermarked voice stream. The solution's general idea is presented in Fig. 1.

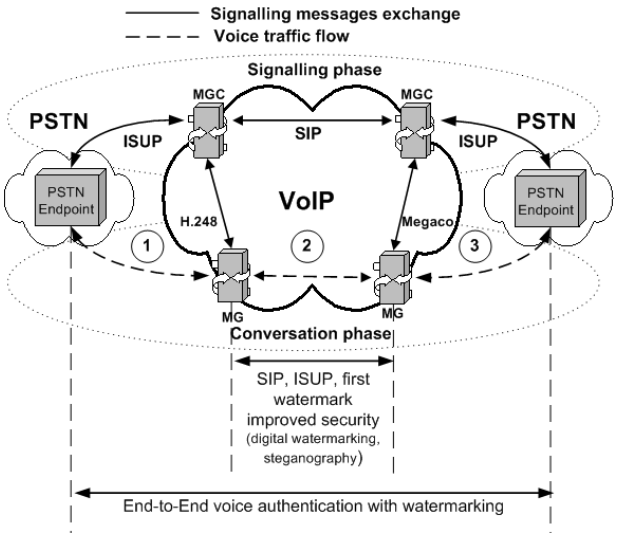

Fig. 1 The proposed solution's general operation idea

We will reference and describe the above figure in the next sections. In next two sections we will describe the proposed solution in details for IP, then PSTN part of the PSTN-IP-PSTN scenario.

4. The IP part of the PSTN-IP-PSTN scenario

In this section we will focus only on the part of the Fig.1 that is marked with ②. We will demonstrate the IP part of the solution based on the SIP protocol [6] and its extension SIP-T [7], because it is now one of the most popular signalling protocol for IP Telephony. SIP-T (SIP for Telephones) is an extension to SIP protocol that allows it to be used for the ISUP call setup between the SS7-based public switched telephone and the SIP-based IP telephony networks. SIP-T carries an ISUP message payload in the body of a SIP message. Transporting ISUP in SIP bodies (in public IP networks) may provide opportunities for the abuse, fraud, and privacy concerns, especially when SIP-T requests can be generated, inspected or modified during the travel in IP environment. The standard proposes that ISUP MIME bodies should be secured (preferably with S/MIME) to alleviate this concern. But MIME (as well as S/MIME) has certain disadvantages for real-time services, as stated in [5, 11].

In this paper MGC (Media Gateway Controller) and MG (Media Gateway) concepts are also used for the VoIP-PSTN interworking. MG converts media provided in one type of the network to the format required by another one. MGC controls the parts of the call state that pertain to connection control for media channels in MGs. Interactions between MGC and MG are described in IETF and ITU joint standard H.248/Megaco [8].

Our solution will be showed on the PSTN-IP-PSTN scenario characteristic for SIP protocol that is defined in [7]: SIP bridging. It means that the IP network (with SIP as a signalling protocol for VoIP) will be used only as a transport network (IP backbone); both caller and callee are in PSTN. This is a popular scenario for today's VoIP providers, which is also called *toll-by-passing*.

4.1 Digital watermarking scheme modifications and token generation

Let us present the watermarking scheme (originally proposed in [1]) that will be also used here. Every Media Gateway (MG) is equipped with the functional block called Pre-processing Stage (PPS), see Fig. 2, which is responsible for preparing data before the watermark embedding stage. Its elements operate as follows:

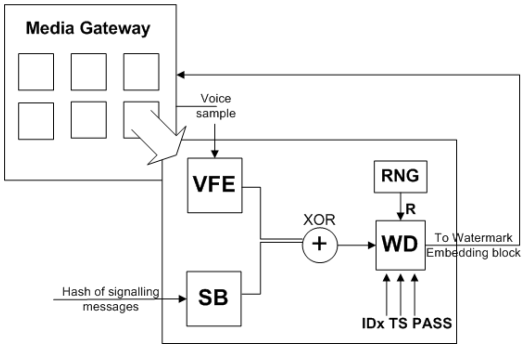

Fig. 2. Pre-processing stage block (PPS) in Media Gateway (MG)

- **SB** (Signalling Message Hash Buffer): stores the hashes of the signalling messages (or those fields that are not changed during transmission) from the first phase of the call provided from MGC (for ISUP and SIP protocols),
- **VFE** (Voice Feature Extractor): provides characteristic features (VF) of the original voice that is sent from PSTN network. Afterwards, a hash function can be optionally performed on this value,
- **RNG** (Randomizer): generates a random number R. It will be used to provide a unique set of the data for every embedded watermark,
- **WD** (Watermarking Data): in this block the input data is concatenated with R, IDX (unique, global identifier of one party of the connection), PASS (the password known to the both parties) and TS (time stamp).

The embedded watermark, formed as described above, we will call a **token**. At the entrance and at the end of the IP network part of PSTN-IP-PSTN scenario, the received

token will be compared with a locally calculated, appropriate one. Fig. 3 shows, how the algorithm works for the SIP protocol with SIP-T extension (to protect ISUP messages that are contained within SIP messages). It is SIP bridging scenario: the caller and callee are in PSTN and their communication is sent through the IP network (MG sends RTP packets to the other Media Gateway).

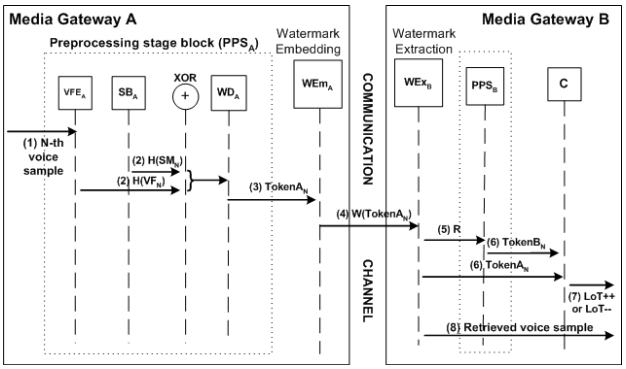

Fig. 3 Proposed authentication solution operation for MG-MG connection

In this situation, the values of the tokens A_N and B_N are:

$$\mathbf{TokenA_N = TokenB_N = H}\left(\left(\mathbf{H(SM_N) \oplus H(VF_N)} \right) \| \begin{pmatrix} \mathbf{TS} \\ \mathbf{PASS} \\ \mathbf{ID_A} \end{pmatrix} \| \mathbf{R} \right) \| \mathbf{R}$$

PPS_B block is analogous to PPS_A in functioning. Additionally, we assume that in the initial signalling phase some of the signalling messages, for both signalling protocols: ISUP and SIP (SM_N means the N-th Signalling Message) were exchanged (and their hashes are stored in SB block). In the second phase they are verified. **H** stands for the hash function and **W** for embedding of the digital watermark into audio. The algorithm works as follows:

- When the conversation begins, the first voice sample enters VFE_A block, the feature of the voice sample (VF_N) is extracted (for the data integrity) and then the hash function is (optionally) performed on the result,
- The values from SB_A and VFE_A are then XORed (they have the same length). Afterwards, the result is sent to WD_A block, in which **TokenA** is created together with the other parameters like: the randomizer value (R), shared password (PASS), global identifier of A (ID_A) and, optionally, the time stamp (TS),
- TokenA is sent to the watermark embedding function and the information, that it contains, is embedded into the caller's voice. Then, the data stream is formed and sent through the communication channel,
- Before the voice from MGA can be processed by MGB, the watermark is extracted and sent to the comparator (C) on the receiver's side,
- From the extracted token, the randomizer value (R) is sent to the analogous PPS_B block. In this block some pre-processing had taken place. Then it computes **TokenB**. It should be equal to TokenA, if the transmission had not been tampered. The result is sent to the comparator (C),
- In C both token values are compared,
- If TokenA=TokenB, the special parameter **LoT** (Level of Trust) is increased (it reflects a number of the correct tokens received). Otherwise its value decreases. Then, depending on the LoT value the decision is made (by MGC), whether the call should be continued or broken down. Computation the LoT parameter is described in details in [1, 2],
- If MGC decides that the call can be continued, the voice sample finally can be converted to PSTN format and directed to callee.

Thus, the authentication and integrity processes depend on exchanging the security tokens and their comparison with the ones locally calculated at the receiver. It is essential to deal properly with every signalling messages exchanged during the connection. Those messages should not influence the call until they are authenticated and their integrity is verified. Authentication of the messages, which are used to terminate VoIP conversations, have to be treated the same as normal messages that come during the call. Normally, the media channels are terminated, upon receiving this message. In our scheme it is vital to retain RTP flow until those messages are authenticated.

As we mentioned earlier, MGC and MG are important for PSTN-VoIP interconnections. H.248/Megaco protocol was created for this purpose. Based on [8] we propose that MG uses *Notify* message to inform MGC that calculated and received tokens are the same or not. MGC should make a decision to continue or break a connection. Almost all the calculations are passed to MG, but MGC must perform a hash function on the signalling messages (or on parts of them) that were exchanged during the first phase of the call (and later). Then, MGC sends those hash values to be stored in MG.

4.2 Network steganography to distinguish security parameters

Besides using digital watermarking in the IP part we will use network steganography to be able to send additional data as described in details in [2, 3]. In this way we will create a multipurpose covert channel. For IP network part we will use it generally for post authentication (as described in [3]). The general protocol overview is presented in the Fig. 4:

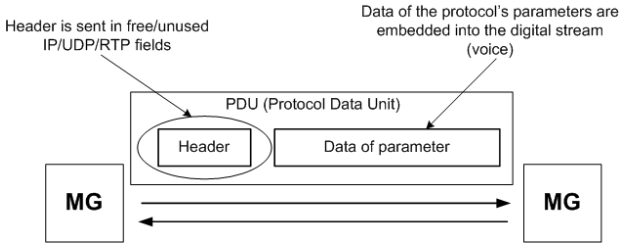

Fig. 4. Using network steganography to distinguish parameters to send

The size of the header of the PDU (Protocol Data Unit), showed in the Fig. 4, is small. In [3] it contained 6 bits that were entered in the unused IP/UDP/RTP fields in the certain way known to both sides of the communication. Moreover, PDU can have one of two the payload types: security or informational. How the informational parameters can be used freely to carry any data that is needed in the communication process, e.g., they can be used to alternate RTCP protocol functionality as we showed in [2]. But more important for proposed application are the security payloads. Security payload means that PDU contains certain authentication and/or integrity information that should be verified after its extraction from the voice. Two kinds of the security payloads are available, first is used to provide authentication and integrity of the voice, its source and signalling protocol messages (e.g. a token, as described in Section 4.1). The second's role is to provide post-authentication for the protocol parameters that were send earlier (both, security and informational). Such a mechanism provides greater security of the whole digital stream and the transmission. The general idea of its calculation is presented in Fig. 5.

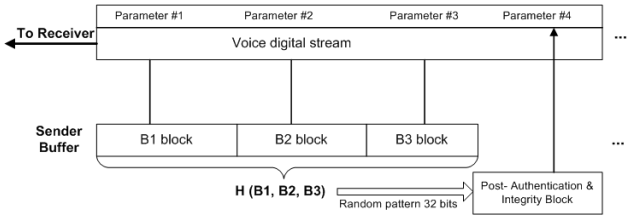

Fig. 5. Example of the post-authentication of the transmitted parameters

As it is presented in the Fig. 5 there is a chaining among the transmitted parameters. Every *n*-th parameter is used for post authentication of *n*-1 parameters that were sent earlier (for Fig. 5. *n*=4).

5. The PSTN part of the PSTN-IP-PSTN scenario

To fulfil the requirements for the proposed solution, the special enhanced PSTN endpoint must be used. It must be capable of the embedding/extraction watermark. We do not dictate audio watermarking technique that will be used in the PSTN endpoint to embed watermark into voice, although, it can be a digital or an analogue watermarking scheme. The important thing is that the watermark that is created in the voice must survive AD/DA operation and embedding of the different watermarks (the watermarks should not overwrite). For the PSTN endpoint the two modes are available:

- It can embed a watermark using an analogue watermarking technique. In this case the watermark must be characteristic for a particular user. The watermark created must also reflect special features of the conversation sent, as it was shown in the case of the IP part of PSTN-IP-PSTN connection in the Section 4.1.

- It can convert the analogue voice into a digital domain, embed an analogous watermark as presented in the Section 4.1, and then convert it again to the analogue signal and transmit through PSTN network.

At the endpoint, the watermark is embedded and it is verified at the other end of PSTN-IP-PSTN connection (① and ③ in Fig.1). During the retrieving process, the other PSTN user verifies only the watermark embedded by the caller. Some other data (added during the travelling through the IP network part) is ignored.

6. Security services for the PSTN-IP-PSTN scenario

The proposed solution, shown in Fig. 1, that uses watermarking technique and network steganography, significantly improves security for PSTN-IP-PSTN scenario. In the PSTN part it provides (① and ③ in Fig.1):

- **Authentication of the data source** (one can be sure of the identity of the caller),
- **Data authentication and integrity** (one can be sure that the audio comes from the caller and it has not been tampered).

In the IP network part (in Fig. 1 marked with ②) it provides, besides ones mentioned above, the following security services:

- **Authentication of the signalling messages for SIP and ISUP** (one can prove that the caller is the source of the signalling messages that were exchanged during the signalling phase of the call),
- **Signalling messages integrity** (one knows that the signalling messages were not modified during the transmission through the communication channel).

Additionally, security provided in the IP part is enhanced with the use of the post authentication parameter as described in the Section 4.2.

As we see, with the use of the proposed mechanism we can provide certain security services for the voice communication between a VoIP system and PSTN. Moreover, security services provided in the PSTN part are guaranteed in the end-to-end manner.

7. Conclusions

The proposed, lightweight PSTN-VoIP secure cooperation solution defines a useful mechanism that provides authentication and integrity of the voice traffic (both conversation and signalling messages) along the communication path. It takes advantage of the two information-hiding techniques: watermarking and network steganography, so it does not consume any user bandwidth. By using those two techniques it utilizes covert channels to exchange certain security information (but data sent may not be limited for security purposes) between calling parties. Additionally there is a mechanism which chains the data blocks together which result in their greater security.

As showed in this paper, it is suitable especially for the PSTN-IP-PSTN scenario, as it adjusts the IP network part (which we assumed is potentially dangerous) security to the PSTN security level.

We are able to verify authentication of the both: media streams as well as the signalling protocols used (ISUP, SIP). In this way we gain one of the first security mechanisms that works in the heterogeneous environment for the voice communication. Unlike existing security mechanisms like VPNs, it is characterized by low bandwidth consummation, low processing power requirements and provides end-to-end authentication and integrity of the voice and signalling messages. That is why this solution is able to offer reasonable trade-off between providing security and performance even for a low-bandwidth environments.

References

[1] W. Mazurczyk, Z. Kotulski. New VoIP traffic security scheme with digital watermarking. In: Proceedings of SafeComp 2006, Lecture Notes in Computer Science 4166, pp. 170 - 181, Springer-Verlag, Heidelberg 2006.

[2] W.Mazurczyk, Z.Kotulski, New security and control protocol for VoIP based on steganography and digital watermarking, Annales UMCS, Informatica, AI 5 (2006), pp. 417-426. ISNN 1732-1360..

[3] W. Mazurczyk, Z. Kotulski. Covert channel for improving VoIP security. In: Proceedings of The 13th International Multi-Conference on Advanced Computer Systems ACS 2006, Vol. 1, pp. 361-370, Międzyzdroje 18-20 October, ISBN 83-87362-75-1.

[4] S. Yuan, S. Huss, "Audio Watermarking Algorithm for Real-time Speech Integrity and Authentication", Proc. Multimedia and Security Workshop, Magdeburg, Germany (2004) pp. 220 – 226.

[5] D.R.Kuhn, T.J.Walsh, S.Fries, "Security Considerations for Voice Over IP Systems", NIST Special Publication 800-58, January 2005.

[6] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, RFC 3261 "SIP: Session Initiation Protocol", IETF, June 2002.

[7] A. Vemuri, J. Peterson, RFC 3372 "Session Initiation Protocol for Telephones (SIP-T): Context and Architectures", IETF, September 2002

[8] F. Cuervo, N. Greene, A. Rayhan, C. Huitema, B. Rosen, J. Segers, RFC 3015 "Megaco Protocol Version 1.0", IETF, November 2000.

[9] Skype: http://www.skype.com

[10] R. Barbieri, D. Bruschi, E Rosti, "Voice over IPsec: Analysis and Solutions". Proceedings of the 18th Annual Computer Security Applications Conference, 2002.

[11] D. Geneiatakis, G. Kambourakis, T. Dagiuklas, C. Lambrinoudakis, S. Gritzalis, "SIP Security Mechanisms: A state-of-the-art review", in the Proceedings of the Fifth International Network Conference (INC 2005) p. 147-155, July 2005, Samos, Greece.

Wojciech Mazurczyk received the B.S. and M.S. degrees in Telecommunication from Warsaw University of Technology, Faculty of Electronics and Information, Institute of Telecommunication in 2003 and 2004, respectively. Currently he is a Research Assistant at Warsaw University of Technology and is finishing his thesis about evaluating information hiding techniques (digital watermarking and steganography) for improving Voice over IP service security. Member of Network Security Group at Department of Electronics and Information Technology of Warsaw University of Technology, Poland.

Zbigniew Kotulski received his M.Sc. in applied mathematics from Warsaw University of Technology and Ph.D. and D.Sc. Degrees from Institute of Fundamental Technological Research of the Polish Academy of Sciences. He is currently professor at IFTR PAS and professor and head of Security Research Group at Department of Electronics and Information Technology of Warsaw University of Technology, Poland.