Intrusion detection systems for IP telephony networks

Martin Steinebach*, Frank Siebenhaar[x], Jana Dittmann[#],
Utz Roedig[+], Ralf Ackermann[+], Christian Neubauer[x]
*Fraunhofer Institute IPSI, Dolivostr.15, 64293 Darmstadt, Germany
martin.steinebach@ipsi.fraunhofer.de
[x]Fraunhofer Institute IIS, Am Weichselgarten 3, 91058 Erlangen, Germany
sbn@iis.fhg.de, neu@iis.fhg.de
[#]Platanista GmbH, jana.dittmann@platanista.de
[+]KOM Multimedia Communiucations, Merckstr. 25, 64283 Darmstadt, Germany
Utz.Roedig@KOM.tu-darmstadt.de , Ralf.Ackermann@KOM.tu-darmstadt.de

Summary

Intrusion detection systems (IDS) provide security for network systems. They are used in computer networks to detect violations against security policies or unusual events that could lead towards a security thread. Telephone networks based on the internet protocol (IP) called IP telephony (IPT) are a recent development in network usage and will become a common application in the next years as they can provide integrated services based on telephony communications.
We identify IPT security demands as well as risks and analyze the possibility of adding IDS concepts to IPT systems. Digital audio watermarking is a technology able to provide different security aspects in this context. The combination of classic IDS and audio watermarking leads to new and promising way of user and data authentication and monitoring of calls.

1. Background and Motivation

In this section we briefly introduce the IP telephony scenario and the technology used for IP telephony Furthermore we discuss the security risks of IP technology inclusion.

1.1 Pre-IP telephony networks

Telecommunication is often an area with a high security demand. The Internet on the other hand is a risky and insecure environment. Therefore methods for securing IPT are necessary to ensure user acceptance of this new service. For existing telephony systems fraud detection is an established service. In the IPT network, similar services are required and IDS often functions to protocol connections and to discriminate between correct use and fraud.
Figure 1 shows a telecommunication network including IP-telephony terminals (T) and standard telephones (image). Telephony security is based on the identification of the telephone used for communication. “Identification by wire” provides a connection between the possibility of using the network and a correct identification of calling and called party. A manipulation can only take place in the physical domain. Another telephone can be attached at the telephone wire to use the same caller ID as the original owner of the number.

1.2 IP Telephony

IPT scenario networks are more complex due to the connection to the classical telecommunication networks. Computers or terminals (which are basically simple computers) are used instead of telephones. Requirements are a multimedia capability for recording and playback of audio data and IP access. Different protocols like H.323 or SIP can be applied to achieve connectivity with other calling partners. Audio codecs are used for AD/DA conversion and compression of the audio data resulting in a trade-off between sound quality and bandwidth requirements.

Steinebach, Siebenhaar, Neubauer, Ackermann, Roedig, Dittmann; Intrusion Detection Systems for IP Telephony Networks, Real time intrusion detection symposium, Estoril, Portugal, 2002 CDR, ISBN 92-837-0032-5, Meeting Proceedings RTO-MP-101, 2003

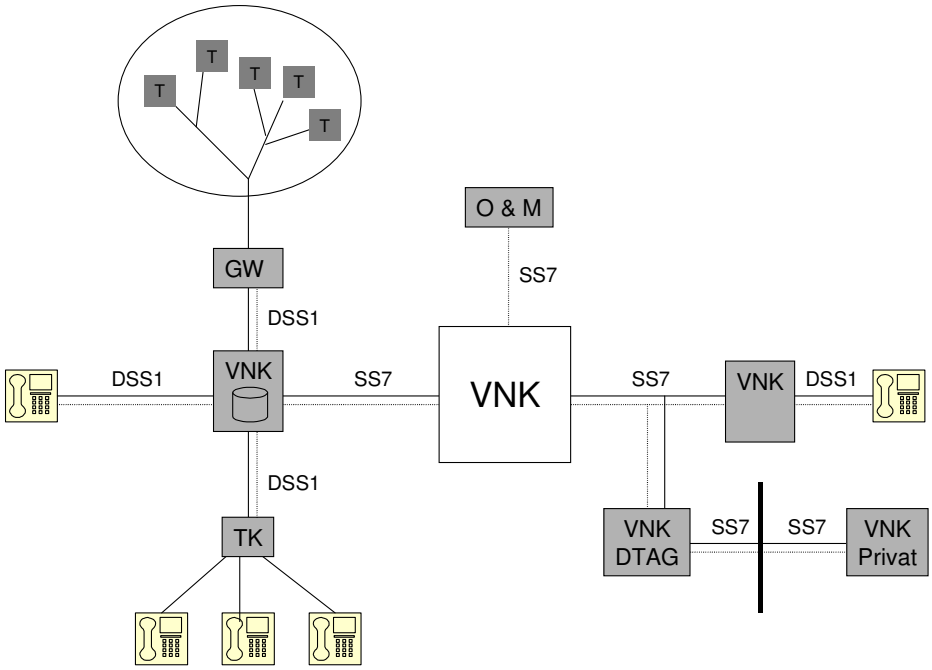

Figure 1: Mixed telecommunication network

1.3 IP-telephony security issues

By using a computer for communication, especially trough a network, identification by wire is lost. Instead a identification by computer or by user ID can take place. Both can be attacked as they are based on transmitted data. Computer networks offer numerous points of attack. Tools for spoofing can be used to act as another calling party. In [ASRS01] security issues are discussed in more detail.
Intrusion detection is a widely accepted technology to address security problems in computer networks (see e.g. [SBD+91]). But, compared to most computer networks secured by an intrusion detection system (IDS), a telephony network is much more complex:

- IP telephony switches from computer to telephony networks and/or vice versa
- network connections can be international, using multinational servers for connection
- mobility is an important issue for IP telephony, a caller or terminal may move from one network to another, taking rights and policies with him / it.

Therefore new, distributed concepts are necessary to achieve IPT intrusion detection. In our paper we introduce an approach by combining IDS concepts with traditional telecommunication based methods. Furthermore we integrate digital watermarking into our IDS-concept. In an IPT network, spoken words are transmitted as sampled audio data. Digital watermarking technologies can provide solutions for authenticity and integrity risks. Digital watermarks can be transparently added to existing voice over IP systems and are designed to survive the applied compression codecs.

2. Digital watermarking

Digital Watermarking is a powerful technology capable of solving important practical security problems like authentication for copyright protection. Watermarking techniques usually used for digital imagery and now also used for audio and 3D-models are relatively new and are growing at an exponential rate. Well over 90% of all publications in this field have been published in the last 7 years. It is a highly multidisciplinary field that combines image and signal processing with cryptography, communication theory, coding theory, signal compression, and the theory of visual perception. Interest in this field has recently increased because of the wide spectrum of applications it addresses.

As an introduction, we discuss the main watermarking parameters. The most important properties of digital watermarking techniques are robustness, security, imperceptibility/ transparency, complexity, capacity and possibility of verification [Dit00], [CMB2002].

- **Robustness** describes if the watermark can be reliably detected after media operations. We emphasize that robustness does not include attacks on the embedding scheme that are based on the knowledge of the embedding algorithm or on the availability of the detector function. Robustness means resistance to "blind", non-targeted modifications, or common media operations.
- **Security** describes if the embedded watermarking information cannot be removed beyond reliable detection by targeted attacks based on a full knowledge of the embedding algorithm and the detector, except the key, and the knowledge of at least one watermarked data. The security aspect also addresses the false positive detection rates.
- **Transparency** is based on the properties of the human visual system or the human auditory system. A transparent watermark causes no artifacts or quality loss.
- **Complexity** describes the effort and computational time we need for watermark embedding and retrieval. This parameter is essential if we have real-time applications. Another aspect addresses if we need the original data in the retrieval process or not. Here we distinguish also between non-blind and blind watermarking schemes which influences the complexity.
- **Capacity** describes how many information bits can be embedded. It addresses also the possibility of embedding multiple watermarks in one media in parallel.
- The **verification** procedure describes if we have a private verification like private key functions or a public verification possibility like the public key algorithms in cryptography.

The optimization of the parameters is mutually competitive if we want to embed a large message, we cannot require at the same time large robustness. A reasonable compromise is always a necessity.
We proposed a media independent classification scheme as a basis for quality evaluation in [Dit00]. It is oriented on the application areas where watermarking techniques can be used to meet the needs of the users. Usually existing watermarking techniques can be used in several applications but in each application it is hard to fulfill all quality demands. We find the following watermarking classes based on application areas for digital watermarking:

- **Authentication or Copyright Watermark:** Ensuring copyright protection by watermarking the data with an owner or producer identification
- **Fingerprint Watermark:** Ensuring copyright protection by watermarking the data with customer identifications to track and trace legal or illegal copies
- **Copy Control or Broadcast Watermark:** Ensuring copyrights with customer rights protocols, for example for copy or receipt control
- **Annotation Watermark:** Ensuring copyrights by annotations or capturing of the media data, this kind of watermark is also used to embed descriptions of the value or content of the data
- **Integrity Watermark:** beside the authentication of the author or producer we want to ensure integrity of the date and recognize manipulations

For the IP-telephony intrusion detection scenario, the following applications are of special interest:

- **Authentication watermarking** can ensure the correct identity of calling parties or terminals. A watermark embedded into the voice stream will be harder to attack then a authentication sent by a network protocol. Both calling parties and the IDS can use the embedded watermark to identify the source of the audio stream. This requires a framework for key exchange like a public key infrastructure. Figure 2 shows an example scenario. Here A could use his secret key to encrypt his ID information before it is embedded. Now B who knows who claims to have called him can receive the public key of A and try to decode the embedded information. If this is successful, B can be sure that A has encoded the ID. Current time and data e.g. would be a good candidate for a embedded message, as this would disable replay attacks.
- **Integrity watermarking** can ensure the audio stream has not been tampered. A mechanism in the called terminal and/or in the IDS sensors can identify if a fragile watermark embedded in the calling terminal has been destroyed. Various concepts for fragile watermarking exist, resulting in different possible concepts for integrity protection in this scenario ranging from detection of time gaps to content protection of spoken information.
- **Broadcast watermarking** could be used to identify attackers listening to communications. If a watermark with the ID of the called party is embedded into the audio stream, a IDS sensor

could identify this ID and check if the called party can be found in the network the IDS is monitoring. If this is not the case, the IDS can inform a firewall to block the incoming audio stream and prevent any user inside the network from receiving the audio data,

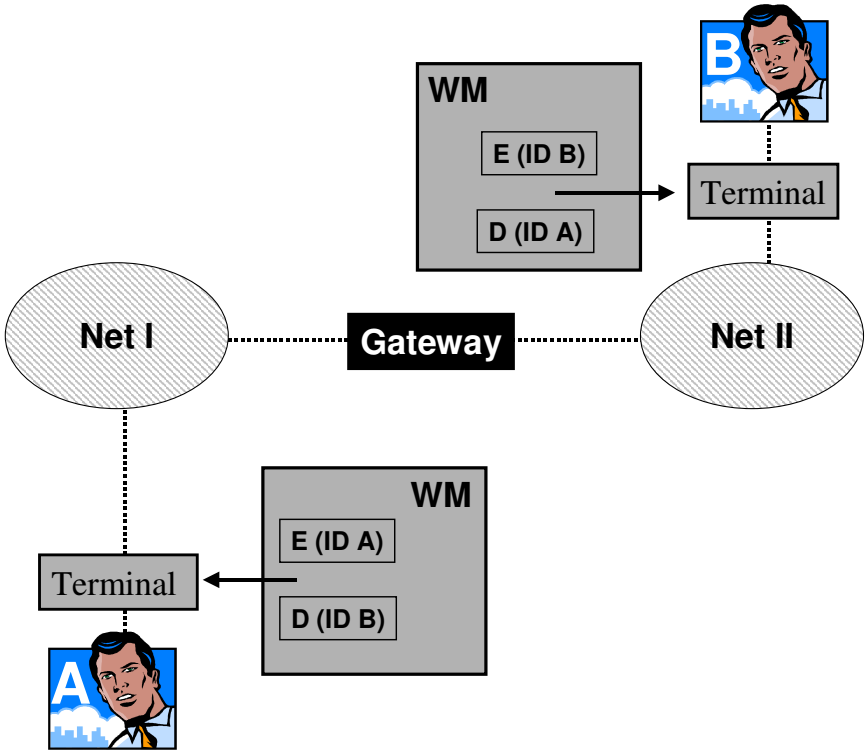

Figure 2: A simple watermarking-based authentications scenario.

2.1 Audio watermarking robustness to telephony

To establish a security concept for IP-telephony based on digital audio watermarking, the applied watermarking algorithms have at least to be robust against the inherent attacks of the scenario:

1. **Low bit rate audio compression:** In voice communication, a strong reduction of the amount of transmitted data is usual. The audio data may be present in high quality with 44,1 kHz and 16 bit at the terminal for the embedding algorithm. For transmission, it will be reduced to e.g. 4 bit and 8 kHz.
2. **Codec changes:** In IP telephony, a compression scheme change may occur while transmission. A gate may recognize a data rate not supported in the local network and change the audio format to another compression scheme.
3. **DA/AD conversion**: Not all communication lines in a telephony scenario will be digital. In some cases a change from digital to analogue audio representation will occur.

(1) has been addressed in our paper “Audio watermarking robustness to lossy compression” [REF]. Here we show that even very low bit rates like mp3 with 30 kbps can be survived if the algorithm has been optimized for high robustness. (3) has been the subject of our paper “Audio watermarking robustness to DA/AD conversions”. Like with lossy compression, a high robustness is possible if a loss of perceived audio quality is accepted at the same time. We show that even analogue transmissions via speaker and microphone can be survived. While a simple change from digital to analogue transmission has not been the subject of our tests, first experiments showed a very high robustness to this attack.

Figure 2 shows an example scenario: An authentication watermark for user A is embedded into the PCM voice data. Now a number of attacks against the robustness of the watermark are applied to the audio stream:

- **Lossy compression** of the PCM stream as discussed above.
- The voice stream can be subject to **packet loss** in the network. Usually in real-time environments it is not possible to resend the audio packets. Therefore they are dropped from the stream and result in gaps compared to the original stream. This will be a challenge to the synchronization capabilities of the watermarking algorithm.
- **Format conversion**: The lossy compression format may be changed at a gateway as described above under "codec changes".
- **Packet losses** can also occur in net II.
- The audio data is **converted to PCM** data. Here usually a change of sample rate and number of bits for sample representation will occur, resulting in an addition of noise to the signal.

Now user B tries to identify the calling party by detection of an authentication watermark in the audio stream A watermarking algorithm suited for the scenario has to survive all the attacks above at the same time or is not robust against the scenario. If the watermark has been destroyed, B will not find A's ID and can not trust the communication.

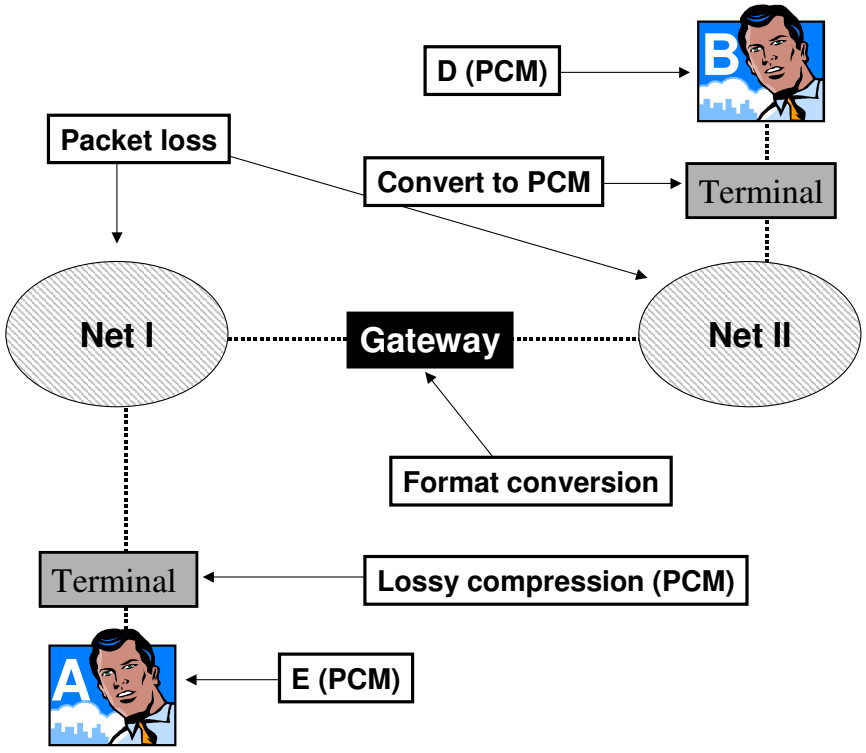

Figure 3: Various attacks against an embedded watermark

2.1.1 Test results

Here we discuss first test results for digital audio watermarking robustness to IP-Telephony based compression schemes. We use our IPSI audio watermarking prototype at different robustness modes and bit rates and checked its robustness against the following compression methods:

- GSM 6.10 at 44,1 and 8 Khz
- mu-Law (CCITT standard G.711)
- Microsoft ACM: Lernout Hauspie CELP 4,8 kBit/s
- 3-bit IMA/DVI ADPCM (5.3:1)

- 4-bit Dialogic ADPCM

The original audio material was encoded in 16 bit, 44,1 kHz , mono. Five different files containing spoken text in English language at various quality levels have been used as examples.

- High payload watermarks with a data rate of ~30 bits per second are robust against GSM 6.10 coding at 44,1 kHz.
- Medium payload watermarks with a data rate of ~7 bits per second are robust against mu-Law, 3-bit IMA/DVI ADPCM and 4-bit Dialogic ADPCM
- Low payload watermarks with a data rate of ~1 bit per second are robust against GSM at 8 kHz.

All watermarking modes have shown to be robust against compression methods higher-payload modes are robust against. The low payload watermark is robust against very high compression rates: GSM at 8 kHz has a compression rate of 1:53 compared to the original audio format.

2.2 Real-time audio watermarking

Besides robustness, low complexity of the watermarking algorithm is a very demanding aspect. By nature, IP-telephony is intended to be a real-time application. This means that the complete processing in the terminals – watermark embedding, audio encoding, and sending of the data via IP in case of the calling party terminal or receiving of the data, audio decoding, and watermark detection in the called party terminal – must be carried out in real-time. Moreover, IP-telephony in general is a two-way communication in which each terminal acts as sender and receiver simultaneously. This intensifies the requirement for low complexity systems.
IP terminals can be computers or also modified telephones which are capable of establishing a connection to the Internet. Considering the first case, the computational complexity for IP-telephony should be low enough that the user can work with other programs at the same time, e.g. a text editor to make some notes about the call. In the latter case, an overall low complexity of all processing steps and, thus, also the watermark embedding/detection algorithm is an important issue in order to minimise the costs for the terminals.

2.2.1 Test results

All prototypic implementations of the watermarking schemes developed at the Fraunhofer IPSI and the Fraunhofer IIS perform several times faster than real-time on a conventional computer system. In the following, the complexity measurement results of the PCM watermarking scheme of the Fraunhofer IIS are presented. This scheme offers a very robust and transparent watermark embedding with a watermark data rate of up to 48 bit/s. The tests were carried out on a state-of-the-art Pentium-4 based PC with 1.6 GHz and 512 MByte of main memory. The complexity numbers given in table 1 are normalised to the duration of the test items (monophonic and stereophonic, 16 bits per sample, 44100 Hz sampling frequency), i.e. 1.0 corresponds to execution of the application in real-time and a number below 1.0 denotes a computation time less than the playing time.

Program	Mono	Stereo
Watermark Embedder	0.070	0.135
Watermark Detector	0.156	0.314

Table 1: Complexity numbers of the Fraunhofer IIS PCM watermarking software evaluated on a Pentium-4 based PC-system.

The table shows that especially in the case of monophonic audio signals – the default signal type for IP-telephony – the watermark embedder performs around 14 times faster and the watermark detector more than six times faster than real-time. However, it has to be kept in mind that this watermarking scheme was developed for high quality audio material. The execution time can be decreased

significantly if the parameters of the watermarking scheme are adapted to the needs of an IP-telephony application. As an example, the bandwidth of the input signals for the watermark embedder is usually limited by a pre-filtering step and, therefore, the calculation of the watermark has only to be performed for this reduced bandwidth.
Thus, it can be stated that the extension of real-time IP-telephony applications with digital watermarking is already possible. Due to the low complexity of the watermarking steps there is still enough remaining computational power for additional processes that can be executed in parallel when applied on modern computer systems. Furthermore, this means that the watermarking functionality can be integrated in specialized hardware terminals using low cost chips.

3 Watermarking based Intrusion Detection

In section 2 we discuss the usability of digital audio watermarking in an IP telephony scenario. As robustness and complexity are both suited for the application, we now describe our concept of watermarking-based intrusion detection for IP telephony as well as other basic security functions based on this approach. In our paper [DKLS2002] we introduce a watermarking-based multimedia firewall. Here a detector is connected with a firewall. Incoming or outgoing media data for which a detector is present is analyzed. If a watermark is successfully detected or (taking the opposite approach, can not be detected properly), different reactions are possible. The most basic one is simply blocking the data. Protocol functions are also described.

3.1 User Authentication

To implement security mechanisms in IP-telephony systems, user authentication is an important requirement. We propose the following concept (see also figure 2):

1. Calling party terminal: Watermark is embedded in audio stream
2. Network: Watermark is transported with the audio data. In case of codec-changes or DA/AD conversion, the watermark stays present.
3. Called party terminal: Watermark is detected

Based on the embedded watermark the called party can now identify the caller. This makes it hard for a third party to spoof a connection as the specific watermarking key is not known. The called party can also detect sudden changes in the embedded identity and e.g. alert the caller or simply block the connection. The security of the embedded watermark can be increased by applying more sophisticated protocol extensions. The embedded caller ID, actual time, called-party ID and other info can be encrypted using a PKI private key and then be retrieved and decrypted using the corresponding public key.

3.2 Data Authentication

Integrity protection can work similarly to the authentication protocol. We have already proposed concepts for audio integrity protection in [DSS2001]. We propose to embed a time code, a content description or both with a robust watermark in the audio stream. At the other end of the connection, the mark is retrieved. Now changes in sequence of audio packages or manipulations of the content can be detected by comparing the audio data with the embedded mark.

3.3 Intrusion Detection System

Like the end systems, an ID-System for IP-Telephony environments is able to use the watermarking information in the audio stream to fulfill its tasks. The resulting advantage for an ID-System using the watermark information in addition (or as a replacement) to the information extracted from the IP-telephony signaling flows is the following.
An IP-Telephony call might be routed through different protocol clouds (e.g. a SIP and a H.323 cloud) using appropriate gateways. To be able to use IDS-Mechanisms that cover both protocol worlds, security related information regarding the call, have to be present and mapped appropriately in both clouds. Therefore, a gateway has to translate security related signaling protocol elements in addition to the signaling elements used for call signaling and call control. Current available gateways do not possess this capability, because a definition for a security related protocol mapping between different IP-Telephony standards is missing. Additionally, if a call is transported via a non IP Telephony based

link (such as a conventional telephone line) parts of the signaling information usually gets lost at the terminating gateways and only the audio data and what is embedded within stays available. By using watermark information in the audio stream, the mentioned gap could be covered. Because the necessary user-information is present in the audio stream this information is present in all protocol-clouds that are involved in a call. This includes calls that are routed between PSTN and IP-Telephones. In such a scenario, an adaptation between both signaling clouds to support security mechanisms would be hard to achieve. Even gateways that perform a transcoding of the audio streams might be used if robust watermarks are used.

An ID-System using the advantage of watermark information needs access to the PK-Infrastructure that is used for the watermarking process of all involved Telephony systems. In addition the ID-System needs the capability to access the audio streams to extract the watermark. Such an ID-System is capable to detect a number of (suspicious) misuse situations. In the following we give a subset of examples that can be used to qualify a call or a set of calls as possible attacks:

1. Calls that are originated or terminated from users that are not permitted to use the service (because it either has no or a false ID for a certain service, e.g. calling abroad).
2. Frequency of calls within a certain time period (e.g. “War dialing”) – those can even be originated from different protocol clouds (both IP Telephony and non IP Telephony)
3. Suspicious call patterns such as a number of parallel calls originated from different locations but using the same unique ID. This might indicate that a certain account has been corrupted.

When comparing the approach with existing security mechanisms (including authentication and protection) it has a number of advantages. The protection scheme and its integration with IDS mechanisms can cover several heterogeneous protocol clouds and call transmission links. The approach can both be used end-to-end as well as only on parts of the (call) link. Systems that are not able to process watermarks do not interfere with systems that are capable to do so.

4 Summary and Conclusion

The audio watermarking algorithms in our tests are robust against compression codecs typically applied in IP-Telephony. With high compression rates, the bit rate of the algorithms has to drop to provide sufficient robustness. At compression rates of above 50:1, we can still embed ~1 bit per second. Therefore we see audio watermarking as a powerful new technology to improve IP-telephony security. We show how audio watermarking and IDS can be combined to solve open security problems in IP Telephony .

One first approach could be user authentication by embedding user Ids into the audio streams. This also leads to the possibility of new intrusion detection methods based on monitoring the embedded IDs. Current challenges are key management for watermark embedding and detection and computational power for watermarking-enabled IDS sensors.

Later concepts could also include embedding of time codes to disable replay attacks based on recorded voice messages and for detecting gaps in the audio stream. Integrity protection be embedding content-fragile watermarks is also on interesting perspective, but the required data rate combined with a high required robustness makes this a future research topic.

5 Future Work

To develop a watermarking-based intrusion detection system for IP-Telephony and similar applications, a number of improvements are necessary in audio watermarking and framework domain.

Payload and reliability regarding errors have to be improved if one requires integrity protection besides authentication. Our current research shows that content integrity protection for audio watermarking is only possible by applying checksums for feature descriptors.

Our multimedia firewall prototype based on watermark detectors can be seen as a proof of concept. Future research has to improve processing speed of the firewalls’ watermark detector to reduce delay and required buffer memory.

At the Transmark project we are working on improvements in real-time audio and video watermark embedding and detection. A lower computational cost for both embedding and retrieval enables a broader analysis of media streams as a single IDS will be capable of monitoring more connections. At the same time, the requirements for the terminals embedding the watermark in real-time will be reduced.

To summarize our future activities, we will improve watermarking technology and integrate it into existing frameworks for new media-based security systems.

Acknowledgments

Parts of this research have been funded by T-Nova and the BMBF project TransMark.

Literature

[ASRS01] Ralf Ackermann, Markus Schumacher, Utz Roedig, Ralf Steinmetz: Vulnerabilities and Security Limitations of current IP Telephony Systems; Proceedings of Communication and Multmedia Security 2001, Kluver, 2001

[CMB2002] I. Cox, M. Miller, J. Bloom, Digital Watermarking, 2002, Academic Press, San Diego, USA, ISBN 1-55860-714-5

[Ditt00] Dittmann, Jana: Digitale Wasserzeichen, Springer Verlag, ISBN 3 - 540 - 66661 - 3, 2000

[DKLS2002] Jana Dittmann, Stephan Klink, Andreas Lang, Martin Steinebach,"Wasserzeichenunterstützende Firewalls", Enterprise Security: Grundlagen, Strategien, Anwendungen, Realisierungen, Patrick Horster (Hrsg.), it Verlag für Informationstechnik GmbH, Höhenkirchen, pp. 246 - 257, ISBN 3-936052-03-4,2002.

[DSS2001] Dittmann, Jana; Steinebach, Martin; Steinmetz, Ralf (2001). Merkmale digitaler Audiodaten zur Verwendung in inhaltsfragilen digitalen Wasserzeichen. In: Verlässliche IT-Systeme 2001, Sicherheit in komplexen IT-Infrastrukturen, Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig/Wiesbaden, pp. 193 - 208, ISBN 3-528-05782-3, 2001..

[RAS01] Utz Roedig, Ralf Ackermann, Ralf Steinmetz. Sicherheit von IP- Telephonie-Szenarien, Februar 2001.

[RATWS00] Utz Roedig, Ralf Ackermann, Marc Tresse, Lars Wolf, Ralf Steinmetz. Verbesserte Systemsicherheit durch Kombination von IDS und Firewall. In Systemsicherheit, March 2000

[SBD+91] S. Snapp, J. Brentano, G. Dias, T. Goan, T. Grance, et al.; A System for Distributed Instrusion Detection, Proc. of the 14th Department of Energy Computer Security Group Conference, May 1991, pp.(17)25-(17)45.