Elektrotechnik & Informationstechnik (2010) 127/5: 116–120. DOI 10.1007/s00502-010-0730-5

IMS security and what we should learn from the Internet

O. Jung, A. Berger, M. Hirschbichler, I. Gojmerac OVE, H. Lippitsch, M. Tscherwenka, K. Umschaden

The IP Multimedia Subsystem (IMS) was developed by a common initiative of the standardization bodies 3GPP and ETSI TISPAN with the aim to provide a platform for fixed and mobile network operators that supports future communication services. In this article we give an overview about the implications that security threats already known from the Internet have on IMS. We analyze the security requirements of an IMS based network architecture and describe possible security solutions. Finally, we argue that for a secure IMS environment extensive traffic monitoring is required in order to provide IMS-based services in a secure manner.

Keywords: IMS; NGN; DoS; SPIT/SPAM; IDS; SIP; botnets; QoS

IMS-Sicherheit und was wir vom Internet lernen sollten.

Das IP Multimedia Subsystem (IMS) stellt eine gemeinsame Anstrengung der Standardisierungsgremien für Festnetz- und Mobilkommunikation dar, deren Ziele sowohl die Entwicklung einer neuen Architektur für Betreibernetze als auch die Unterstützung zukünftiger Dienste sind. In diesem Artikel geben wir einen Überblick über die Implikationen der aus dem Internet bekannten Sicherheitsproblematiken. Wir analysieren die Sicherheitsanforderungen einer IMS-basierten Netzarchitektur und zeigen mögliche Lösungsansätze auf. Wir kommen zu dem Schluss, dass für eine sichere IMS-Umgebung ein über die Standardisierung hinausgehendes Maß an Netzüberwachung unbedingt notwendig ist, damit die grundlegende Funktionalität sicher gewährleistet werden kann.

Schlüsselwörter: IMS; NGN; DoS; SPIT/SPAM; IDS; SIP; Botnets; QoS

Received January 29, 2010, accepted April 1, 2010

1. Introduction

The IMS was originally defined by 3 GPP to support the convergence of data, voice and other services over an IP-based infrastructure for mobile networks. The fundamental motivation behind is the understanding that operators need to fill the gap between the traditional telecommunications services like voice and Internet technology. The ETSI TISPAN NGN initiative extends IMS to support fixed-line access networks (*ETSI ES 282 001 v3.4.1., 2009*), thus IMS is today the dominating approach towards access-independent Next Generation Networks (NGN). It is designed to be an extremely flexible next-generation platform by which operators are enabled to provide new and innovative services. In addition, the NGN is expected to reduce operational expenditure of operators and service providers. To support these goals, the IMS platform offers versatile and generic service enablers, such as mechanisms for session negotiation and management, Quality of Service (QoS), mobility, and security.

The IMS is mainly based on IETF protocols, such as IP, DNS, HTTP, and SIP. However, these protocols were designed to offer maximum flexibility and to their users without considering security issues adequately, a strategy that has the potential to backfire in IMS scenarios. Furthermore, they come with security problems that are already well-known from the Internet and which are now inherited by the IMS. These known problems need to be anticipated now, *before* the widespread deployment of IMS installations, in order to ensure that appropriate protection mechanisms are in place when IMS goes live. In this article we introduce the most pressing problems and point at possible mitigation strategies.

2. The IMS architecture

In order to understand the security issues that arise with the introduction of the IMS we first describe the IMS architecture and its security features as defined by 3GPP in (*3GPP, 2009b*). The foundation for many security mechanisms is the proper authentication of users and terminals. Due to strong user authentication in the IMS, operators can reliably grant privileges to their users. They are also enabled to offer services such as Single-Sign-On based on successful authentication. This guaranteed identity is one of the unique features that IMS network operators can offer to their customers.

The core functions of the IMS infrastructure are provided by the *Call Session Control Function* (CSCF) that processes the signaling in the IMS. It is distributed over three functional elements, the *Proxy-CSCF* (P-CSCF), the *Serving-CSCF* (S-CSCF), and the *Interrogating-CSCF* (I-CSCF). The security of the IMS is to a large extent based on secret keys managed by the *Home Subscriber Server* (HSS), which is also responsible for storing user profiles.

The IMS network is divided into security domains, where a security domain is usually corresponding to the network of a single operator. A security domain is a part of the network that is under the control of the same organization and that is subject to the same security policy. Different security domains shall be connected by security gateways that are located at the border of the domain and is responsible for securing IP base communication protocols (*3GPP, 2009a*). Security gateways are the enforcement points for security policies. This includes functions like simple IP filtering and firewall

Jung, Oliver, Dr.-Ing., Berger, Andreas, Dipl.-Ing., Gojmerac, Ivan, Dipl.-Ing. Dr., Forschungszentrum Telekommunikation Wien (FTW), Donau-City-Straße 1, 1220 Wien, Österreich; **Hirschbichler, Michael, Dipl.-Ing.,** Institut für Breitbandkommunikation, Technische Universität Wien, Favoritenstraße 9/388, 1040 Wien, Österreich; **Lippitsch, Hans, Ing.,** Telekom Austria TA AG, Arsenal Objekt 22/506, 1030 Wien, Österreich; **Tscherwenka, Mario**, Kapsch CarrierCom AG, Am Europlatz 5, 1120 Wien, Österreich; **Umschaden, Klaus, Dipl.-Ing. Dr.,** Alcatel-Lucent Austria AG, Scheydgasse 41, 1210 Wien, Österreich (E-mail: jung@FTW.at)

functionality but they also act as Session Border Controller that usually provide a broad range of security related functions like e.g. limitation of connection rates or bandwidth. Additionally, security gateways acts also as IPsec Gateways in order to encrypt inter domain traffic.

Services provided by the IMS are hosted by so-called Application Servers (ASes). An AS can be located in an operator's home network as well as in an external 3rd party network. It is controlled by the IMS core using SIP, but also other protocols like HTTP are used to interface with an AS. In general, the AS is the entity in the IMS where Internet services meet the traditional telecommunications infrastructure.

The IMS mainly uses protocols defined by IETF like the *Session Initiation Protocol* (SIP) (*Rosenberg, et al. 2002*) and DIAMETER to name only the most important ones. The DIAMETER protocol (*Calhoun, 2003*) provides a framework for Authentication, Authorization, and Accounting (AAA) and is amongst others used for the exchange of key material and user data between S-CSCF and HSS. The protocol used for controlling multimedia sessions is SIP. It is a text-based protocol and offers only a small number of message types called methods. However, SIP is an extremely versatile protocol, with a lots of available extensions, and thus capable to support many different applications. IMS signaling using SIP is quite different from Signaling System 7 (SS7) signalling in the Public Switched Telephone Network (PSTN). The main difference is that IMS uses in-band signalling over open interfaces while the PSTN uses a separated network for call signalling. Furthermore and in contrast to SS7, SIP internals are easily accessible and thus available to adversaries.

3. IMS security threats

The preceding section illustrated the complexity of the IMS architecture and showed how classic security issues like authentication and interconnection are addressed. However, these measures aim mostly at security issues as known from e.g. GSM networks, and neglect the Internet heritage of the IMS. Essentially, the IMS is a rather open, IP-based, versatile service platform with powerful signaling mechanisms, which enables the user to communicate directly with the core network. It is this combination of being an umbrella for many different services with typically high availability requirements and large amounts of confidential user information (e.g. personal data like pictures or the user's location) on the one hand, and far less possibilities (as compared to traditional telecommunication networks) to restrict user activity on the other hand, that makes the protection of the IMS challenging.

A close analysis of IMS security shows many similar vulnerabilities in the IMS and the Internet (see (*Berger, Gojmerac, Jung, 2009*) for a detailed survey). Just as the Internet, the IMS provides no standard mechanism to deal with malicious overloading of network resources, i.e. (Distributed) Denial of Service ((D)DoS) attacks. Likewise, there is no specified way to deal with unsolicited communication, as e.g. Spam or, possibly even more annoying, voice Spam (Spam over Internet Telephony (SPIT)). In the following, we outline these key problems of future IMS deployments before we summarize the current state of the art in relevant defensive approaches in the next section.

3.1 Botnets – the root of all evil

Botnets are today considered an enabling technology for most Internet security problems and are thus of major relevance for the IMS. A botnet is basically a large group of coordinated malware instances, that are used for DDoS attacks, Spam distribution, information theft, fraud, and blackmailing. Note that this list can be arbitrarily extended, as it is up to the owner of the botnet (the so-called *botmaster*) which purpose it serves. By introducing the IMS, the operators move away from their current sealed-off core networks with highly restrictive signaling, limited functionality, and rather low-performance user equipment, and consequently botnets become a severe threat to basic telecommunication services like telephony. The expected aggregation of user information in the IMS and its many commercial (and possibly abusable) services make it an even more attractive target.

3.1.1. Why botnets are harmful for the IMS

Botnets usually run through a three-step life cycle: after the initial *infection* of a host with some malicious software, specific *Command and Control* (C&C) mechanisms are used to connect the new bot to the botnet and to receive individual tasks. Finally, these tasks are carried out in the *execution* phase, i.e. the bots start a spamming campaign or flood a target host. Traditional telecommunication networks were immune against these kind of threats, due to simple, not infectable user equipment (i.e. telephones), less complex and much more restricted signaling (e.g. SS7), and a sealed-off core network. In the following we show that these restrictions do not apply to the IMS anymore and that all three phases of a botnet life cycle are feasible within this new environment.

IMS user equipment ranges from fixed line hard phones over mobile handsets and smartphones to standard PCs and is capable to serve rich multimedia services. With increased complexity always comes increased vulnerability, and hence it is not surprising that we see hard phones being hacked[1] and mobile devices being considered increasingly vulnerable (*Ahamad et al., 2008*). Malware infected PCs are without doubt commonplace, anyway. Knowing about the long history of security incidents in the Internet, we can imagine what could happen when we bring together these problems and core telecommunication services. Thus, the infection of a IMS terminal with malware will be clearly possible (see Fig. 1 for an overview of possible botnet infections in an IMS setting).

Infected terminals are themselves a minor risk, unless they are able to communicate with each other. As Berger and Hefeeda show in (*Berger, Hefeeda, 2009*), it turns out that ironically the IMS framework itself is a perfect platform for botnet communication. The efficient and robust coordination of a large number of bots is not trivial, and thus the multitude of existing service enablers like error handling and session control come in handy. However, even when this problem ends up being solved by rigid network monitoring and heavily restricted offered functionality, malware can still use their traditional communication models, as future IMS terminals are naturally expected to be connected to the open Internet. Botnet C&C is thus also going to be feasible.

With these two requirements being satisfied, it comes without surprise that botnets have the means to cause devastating trouble to IMS networks. Distributed, malicious overloading of the CSCF functions can bring the entire infrastructure down. Compromised IMS user equipment enables access to sensitive user information, as IMS terminals are intended to be authenticated to single-sign-on domains and thus have access to a variety of personal data. As a consequence, improved scamming, spamming, and phishing attacks are possible, as the contained data will seem more plausible to the victim user (e.g. a link that seemingly comes from a friend). While in the Internet a botnet is basically an army of 'IP addresses' under control of a malicious user, an IMS botnet is an army of *user identities*, with all their capabilities and permissions.

[1] http://www.scmagazineus.com/exploiting-voip-vulnerabilities-to-steal-confidential-data/article/111091/.

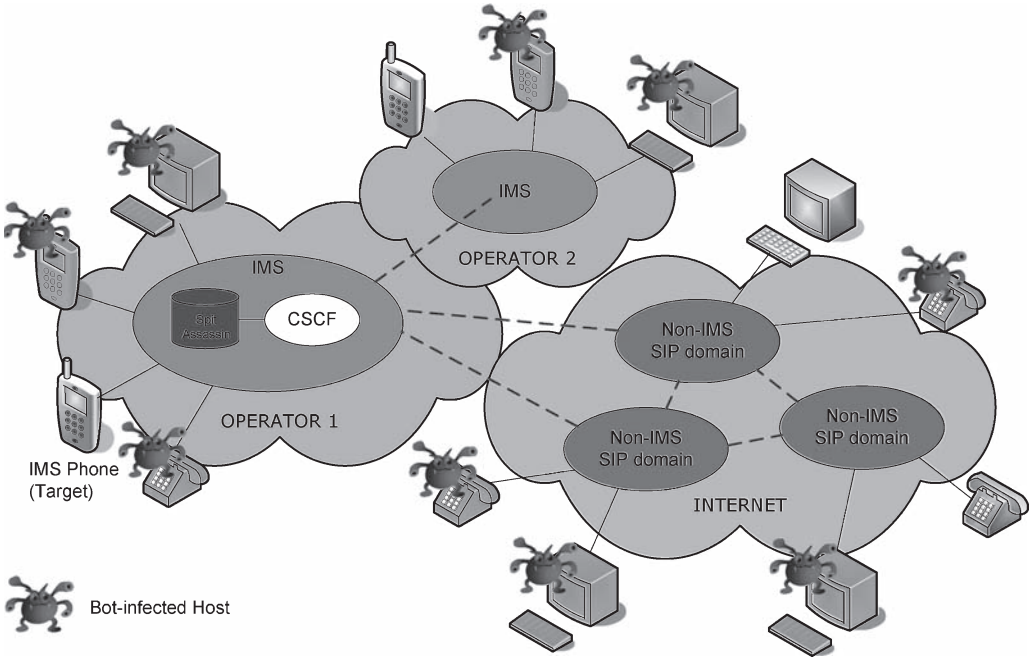

Fig. 1. UC attacks originate from multiple sources and domains

3.2 Unsolicited communication in the NGN

One major threat of botnets is the distributed dissemination of unsolicited communication (UC). Equivalent to Spam in E-Mail communication, SPIT, Spam over Instant Messaging (SPIM) and the presence spam (*Rosenberg, Jennings, 2008*) are a huge annoyance on the one hand, and a major risk for NGN core- and access infrastructure on the other hand. The problem of being called by a call-center-agent or calling robot during nighttime is not only a problem of the called party, but also on the telecommunication providers side, as the provider is putting its reputation at risk.

This threat of UC is well-known from the pre-NGN telecommunication, but with new techniques and possibilities, like video- or textual-messaging, UC in NGN can create a new dimension of annoyant communication. First in-the-wild attacks on VoIP resulting in UC were recorded and analyzed already in 2008 (*Darilion, 2008*).

Fighting the threat of SPIT is not trivial, as the voice-content of a call – in case of SPIT, the promotional message – is played after the called person accepted the call. So, the provider is generally not able to qualify a call as unsolicited in advance, but the called person can – if such a infrastructure is offered by the provider – qualify a call as UC after terminating the call. This input can be used to block the originator to prevent future communication between the distributor of the UC and the called person. This blocking is only successful if the presented identity of the caller is identical to the creator of the UC. If the terminal of the caller is captured by a botnet or the call arrives from another domain which is not under direct or indirect control of the callee's provider, this method is ineffective. We will present possible countermeasures against these types of threat in Sect. 4.2.

Generally, the threat of UC existed already in pre-NGN PSTN/ISDN (*ETSI, 2008*), but the risk of getting UC was tremendously lower because of higher costs, less complex terminals and the strict isolation of the infrastructure against external access. In opposite, in NGN networks, the terminals are more intelligent and the operators offer more interfaces to external hosted application servers, other NGN/IMS operators or Internet-based SIP domains (see 1).

4. Countermeasures

Botnets and unsolicited communication are well-studied in the Internet domain and a multitude of countermeasures have been developed. Currently, none of these approaches are explicitly part of the relevant IMS standards, despite serious security risks. In the following, we give an overview of defensive strategies that should be considered mandatory for secure IMS deployments.

4.1 Botnet mitigation

Effective botnet detection and mitigation requires a thorough understanding of botnet internals. By revisiting their typical three-step life cycle (as introduced in Sect. 3.1), adequate countermeasures can be derived. Based on our experiences with the Internet, we know that malware *infections* of user equipment can hardly be brought completely under control. Although virus scanners and personal firewalls are commonplace today, botnets are flourishing. Many other approaches try to detect and to counter the effects of the *execution* phase. A multitude of intrusion detection and prevention systems are standard equipment today, and can help to minimize the impact of DoS attacks and filter Spam e-mails (see next section). Still, experience showed us that even with these countermeasures deployed, botnets pose a significant threat, as the botnet problem in the Internet is evidently far from being solved.

As a consequence, recently a new field of network monitoring started to be developed, which deals with detection of botnet *Command and Control* traffic (see e.g. (*Strayer et al., 2008*)). This approach has certain advantages: first, the network domain is a controlled environment, as opposed to the user domain. Measurements in this domain cannot be easily manipulated or compromised. And second, the timely detection of C&C traffic helps the network operator to take suitable countermeasures *even before* actual attacks can happen in subsequent botnet execution phases.

To this end, there are a number of challenges to overcome. Most botnets are under continuous development and adapt to countermeasures by changing message rate and syntax or conceal their communication using encryption techniques. Detection of botnet traffic signatures is therefore of only limited use. Consequently,

many recent approaches use machine learning techniques to detect network anomalies caused by botnet communication (e.g. *Livadas et al., 2006*). However, such systems require reliable training records to either learn the normal, uncompromised network behavior or explicitly learn malicious behavior.

4.2 UC countermeasures

In current telecommunication networks, SPIT/UC prevention measures are divided into *legislation and regulation, user authentication* and *contract conditions* (*3GPP: 2009c*). With the distribution of botnets in NGN, these three measures are not satisfactory anymore as the presented caller is not necessarily the initiator of the UC. As a result, the provider must keep an eye on all transactions to identify UC.

These identification activities are divided into three types with increasing level of interaction:

1) Non-intrusive tests: the UC-preventing infrastructure in the NGN analyses the call-signaling to gain information about the UC-probability of an ongoing call-setup. Indicators might be Caller-ID, Source IP, Time-of-Day or, in text-based UC, the payload itself. This information is compared with already identified UC and the conclusion of this comparison can be used to mark a request as UC. Such tests can be done without any interaction with the caller or the callee, but are impeded, when the UC distribution is done by multiple, distributed bot-infected clients.

2) Intrusive test: if the first tests did not result in a significant qualification, the next step is to let the caller and the UC protection mechanism interact directly. This interaction might be a Turing test (*Rosenberg, Jennings, 2008*) to detect, whether a call is created automatically by a bot or manually by a human. Another approach is to block a caller at the first try and prompt him to call later *(greylisting).* This test helps when the calling bot is based on a simple call-and-forget algorithm equivalent to e-mail-Spam.

3) Feedback before, during or after a transaction: the called person can define in advance, by whom he/she wants to be called. This definition can contain the own address book as a whitelist, a maximum SPIT-probability score from the non-intrusive tests or a global reputation list. After a transaction, the called person can add the caller to the white- or to a black-list defining the behaviour of the UC-protection infrastructure the next time the caller creates a call. Botnet-created calls are hence difficult to detect by black- and whitelist-mechanisms, as each call is usually sent from a different terminal and user identity.

The final decision, whether a call should be blocked or not, must fall into the responsibility of the called party. The NGN infrastructure must give the callee the possibility to create settings defining in advance, which calls should be blocked, rerouted or answered on behalf of the UC score. Generally, emergency or priority communications should override the UC preferences.

SpitAssassin

For our prototypical Anti-SPIT implementation *SpitAssassin*[2] we extended the well-known Spam prevention toolkit *SpamAssassin*[3] with SIP analyzing capabilities. Our system successfully analyzed the payload of SPIM and presence spam and added a score-header (e.g.: X-SPIT: 6.3) to the SIP message. As the payload of a SIP INVITE-request just describes an *upcoming* media session with no indication about the content of the pending call, we cannot draw any reliable UC-conclusion from this payload. We propose in (*Hirschbichler et al., 2009*) the use of Spammer IP-blacklists to compare the IPs contained in the SIP messages with the blacklist. Using this lookup, we try to find terminals already known for being infected by a UC-distributing bot and mark all requests transmitted by these terminals with a higher UC score.

SpitAssassin is currently under further development and future extensions will offer the ability to work with decentralized score delivering equipment as proposed in (*3GPP, 2009c*).

[2] SpitAssassin: http://www.spitassassin.org.

[3] SpamAssassin: http://www.spamassassin.org.

5. Conclusion

Extensive network monitoring is mandatory for commercially successful IMS deployments. The responsible standardization bodies have largely neglected this requirement, leaving it up to the individual operators how to implement it. Although industry organisations have started to react and are now developing systems like the previously mentioned SBC and NOCA/GOCAP (a standardized architecture for overload control (*ETSI, 2010*)), these systems are developed in parallel to, and not within, the IMS specifications. Considering in particular the distributed nature of today's attacks, with botnets being the key enabler and core problem, we are convinced that an adequate defensive platform is needed. Standardized and compatible network monitoring, with interconnections between the operators, might be the only way to provide sustained protection. Distributed threats need to be countered by distributed defense, as the current island solutions are not going to be sufficient in the future 'everything-from-everywhere' telecommunication world.

Overall, we are convinced that securing the IMS against Internet-like attacks is going to be extremely challenging. By giving more and more functionality to the user equipment, it becomes more attractive to abuse it. Customers are used to the high availability of basic telecommunication services (i.e., the famous five nines – 99.999%). Operators cannot allow for a degradation here; at the same time they need to take more responsibility for protecting their users. KISS, – Keep It Small and Simple –, is a traditional approach to engineering and especially to security. IMS clearly falls not in this category, and it has yet to prove if it is even so ready for real-world deployments.

Acknowledgement

This work has been supported by the Austrian Government and by the City of Vienna within the competence center program COMET and by the BACCARDI project.

References

Ahamad, M., Amster, D., Barrett, M., Cross, T., Heron, G., Jackson, D., King, J., Lee, W., Naraine, R., Ollmann, G., Ramsey, J., Schmidt, H. A., Traynor, P. (2008): Emerging cyber threats report for 2009. Technical report, Georgia Tech Information Security Center.

Berger, A., Hefeeda, M. (2009): Exploiting SIP for botnet communication. In: Proc. of the 5th Workshop on Secure Network Protocols, Princeton, NJ.

Berger, A., Gojmerac, I., Jung, O. (2009): Internet security meets the IP multimedia subsystem: an overview. Security and Communications Networks.

Calhoun, P., Loughney, J., Guttman, E., Zorn, G., Arkko, J. (2003): Diameter Base Protocol. RFC 3588.

Darilion, K. (2008): Analysis of a VoIP Attack.

ETSI ES 282 001 V3.4.1. (2009): Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN Functional Architecture.

European Telecommunications Standard Institute (ETSI). Feasibility study for prevention of unsolicited communication in the NGN. 2008. TR 187.009 V2.1.1.

European Telecommunications Standard Institute (ETSI). NGN Congestion and Overload Control – Part 2: Core GOCAP and NOCA Entity Behaviours. 2010. ES 283.039–2 V3.1.1.

3rd Generation Partnership Project (3GPP). (2009a): 3 G security; Network Domain Security (NDS); IP network layer security (Release 9). TS 33.210 V9.0.0.

3rd Generation Partnership Project (3GPP) (2009b): IP Multimedia Subsystem (IMS); Stage 2 (Release 9). TS 23.228 V9.2.0.

3rd Generation Partnership Project (3GPP) (2009c): Study of Mechanisms for Protection against Unsolicited Communication for IMS (PUCI). TR 33.937 V9.0.0.

Hirschbichler, M., Egger, C., Pasteka, O., Berger, A. (2009): Using E-Mail SPAM DNS Blacklists for Qualifying the SPAM-over-Internet-Telephony Probability of a SIP Call. In: Third International Conference on Digital Society. ICDS '09: 254–259.

Livadas, C., Walsh, R., Lapsley, D., Strayer, W. T. (2006): Using machine learning techniques to identify botnet traffic. In: Proc. of the 31st IEEE Conference on Local Computer Networks: 967–974.

Rosenberg, J., Jennings, C. (2008): The Session Initiation Protocol (SIP) and Spam. RFC 5039 (Informational).

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., Schooler, E. (2002): SIP: Session Initiation Protocol. RFC 3261.

Strayer, W., Lapsely, D., Walsh, R., Livadas, C. (2008): Botnet detection based on network behavior. In: Botnet Detection, Advances in Information Security: 1–24. Springer-Verlag.

Authors

Oliver Jung
received his diploma in telecommunication engineering from the University of Siegen, Germany. He received his Ph.D. from the same university in 2003. In 2004 he joined Forschungszentrum Telekommunikation Wien, where he currently is involved in national and international research projects as a senior researcher and project manager in the field of telecommunications security. He is member of ISO/IEC JTC1 SC27 (IT security techniques) and ON AG 27 (IT-Sicherheit). His research interests are on NGN security, network security mechanisms, Intrusion Detection Systems, privacy, and identity management.

Andreas Berger
received his M.Sc. degree in Telematik from Graz University of Technology in 2007 and is now working towards his Ph.D. at University of Vienna. He joined Forschungszentrum Telekommunikation Wien in 2007 and currently holds a position as junior researcher. His main research interests are in network security, statistical anomaly detection, and complex network theory.

Michael Hirschbichler
is a Research Associate at Vienna University of Technology, Institute of Broadband Communications. His scientific work focuses on performance evaluation, VoIP security, and, more specifically, IP Multimedia Subsystem security.

Ivan Gojmerac
received his diploma in Electrical Engineering from the University of Zagreb, Croatia, in May 2001, since then he is with the Telecommunications Research Center Vienna (FTW), today as Senior Researcher in the area of packet-based networks. Based on his research in the area of Internet routing, he received his Ph.D. degree with distinction from Vienna University of Technology in 2007; his Ph.D. Thesis on Adaptive Multi-Path Routing for Internet Traffic Engineering has been awarded the GIT-Prize of the Austrian Electrotechnical Association (OVE). His scientific interest further includes 3G networks and the IMS, network simulation and performance evaluation as well as security aspects of future communication networks. Apart from his strategic research, Dr. Gojmerac also actively participates in several application oriented projects at FTW as well as in the EU FP7 research projects PRISM and ETICS.

Hans Lippitsch
joined Telekom Austria TA AG in 2009. His current work focuses on Network Security, and more specifically, VoIP security. Before this, he was with ÖFEG (Österreichische Fernmelde- und Entwicklungsgesellschaft Ges. m.b.H), where he worked as a researcher mainly on the technical realization of innovations in the telecommunications industry.

Mario Tscherwenka
is Customer Solution Manager for Transmission Networks and Security Systems at Kapsch CarrierCom AG. He is responsible for Portfolio & Strategy Planning and Technical Presales for Carrier Grade IP Networks and IT Architectures.

Klaus Umschaden
received his diploma in informatics from Vienna University of Technology in 2002 and a doctorate in 2007. His scientific interests are general security topics and more specifically VoIP security. He currently realizes multimedia projects for the global telecommunications supplier Alcatel-Lucent.

