User-centric, Privacy-Preserving Adaptation for VoIP CAPTCHA Challenges

A. Tasidou[1], P.S. Efraimidis[1], Y. Soupionis[2], L. Mitrou[3,2] and V. Katos[1]

[1]Democritus University of Thrace, Dept. of Electrical and Computer Engineering
[2] Information Security and Critical Infrastructure Protection Research Group, Dept. of Informatics, Athens University of Economics & Business (AUEB)
[3]University of the Aegean, Dept. of Information and Communication Systems Engineering
e-mail: {atasidou, pefraimi, vkatos}@ee.duth.gr; jsoup@aueb.gr; l.mitrou@aegean.gr

Abstract

The effectiveness of CAPTCHA challenges largely depends on being simultaneously easier to solve for humans and harder to solve for bots. In this work we argue that it is possible to enhance the effectiveness of audio CAPTCHA challenges by adapting the challenge to the users' characteristics. We propose a method for achieving this adaptation while protecting users' privacy. Moreover, our approach allows us to address discrimination issues that naturally arise in contemporary audio CAPTCHA challenges. Utilizing modern cryptographic techniques we design a privacy-preserving system, called PrivCAPTCHA, which offers customized CAPTCHA challenges.

Keywords

Privacy Enhancing Technologies, Audio CAPTCHA, SPIT, Discrimination, Legal Aspects of Privacy, Incident Response.

1 Introduction

The Internet Telephony (Voice over IP) is a developing technology that promises a low-cost and high-quality and availability service of multimedia data transmission. Inevitably though, VoIP "inherited" not only these positive features of Internet services, but also some obvious disadvantages (Walsh and Kuhn, 2005). One of the main disadvantages is Spam over Internet Telephony (SPIT) (El Sawda and Urien, 2006), which is the popular expression of Spam in VoIP network environments.

A CAPTCHA (Ahn et al., 2004) is a method that is widely used to counter automated SPAM attacks. The same technique can be used to mitigate SPIT. A CAPTCHA is a Reverse Turing Test where a machine tries to identify whether the incoming session is initiated by a software application or a human. The three major categories of CAPTCHA are a) visual CAPTCHA, where the user tries to recognize characters or words in malformed pictures, b) audio CAPTCHA, where the characters or words to be recognized are in an audio file, and c) logic CAPTCHA, where the user tries to answer specific questions. This paper is focused on audio

CAPTCHA. The reason is that visual CAPTCHA are hard to apply in VoIP systems, mainly due to the limitations of end-user devices. Logic CAPTCHAs are well suited for the VoIP context and are appropriate for adaptive challenges, but one should not neglect the practical difficulties of applying logic CAPTCHAs (Hernández-Castro et al., 2011).

The audio CAPTCHA challenges used today to prevent automated SPIT attacks do not take into account the characteristics of the caller or the callee. Therefore, they need to be easy enough for the average user to solve, making them often easy for automated attacks as well. The fact that these challenges are generic, does not allow for the process to take into consideration the cognitive abilities of human users, while at the same time discriminates against users that have difficulties solving the generic challenges. The ability to use information about the caller or the callee opens up new opportunities for creating more effective and fair CAPTCHA challenges. However, the required information about the caller will probably be sensitive personal information, and thus it is important that a privacy-preserving method of achieving adaptation of CAPTCHA challenges is used.

This work proposes a user-centric, privacy-preserving VoIP CAPTCHA adaptation mechanism , which offers a new paradigm for VoIP CAPTCHA systems aiming at increasing the CAPTCHA effectiveness. We consider this mechanism within the framework of the SPHINX project (Distinguishing Human or Machine with Interactive Media Audio, http://sphinx.vtrip.net). According to the SPHINX architecture, VoIP calls are redirected to an audio CAPTCHA server for evaluation. This work proposes a prototype that could be introduced to the SPHINX architecture. However, there should be no serious difficulties in combining our approach with other audio CAPTCHA systems or even conventional CAPTCHA systems.

1.1 Related work

Existing audio CAPTCHAs are proven more difficult to use for visually impaired than non-visually impaired people (Bigham and Cavender, 2009). For their research they used 162 persons, of whom 89 were visually impaired, and popular website audio CAPTCHA implementations. Their research illustrated that audio CAPTCHAs are difficult to solve. Only 43% of users with visual impairments were able to answer an audio CAPTCHA at the first attempt and only 39% of other users. Moreover, it should be noted that visually impaired users took at least twice as long. Yet nearly half of users (47%) failed to respond correctly to an audio CAPTCHA after 3 attempts. This is a somewhat unexpected result, since one would anticipate that audio CAPTCHA challenges would be more appropriate for visually impaired persons.

E. Bursztein et. al. (Bursztein et al., 2010) conducted an extensive study on the ability of people to solve existing CAPTCHAs, as well. Regarding audio CAPTCHAs, they studied eight of the most popular implementations. The conclusions that emerged from their study were a) the period for listening and solving a CAPTCHA is certainly excessive (averaged over 25 seconds), b) the percentage of users who took second or third attempts, because the previous attempt

was wrong, it exceeded 50%, and c) people who were not native English speakers had major problems in solving the CAPTCHA and therefore the success rate was reduced by more than 20%.

Y. Soupionis and D. Gritzalis (Soupionis and Gritzalis, 2010) classified the audio CAPTCHA attributes, evaluated the current popular audio CAPTCHA implementations and developed a new audio CAPTCHA for VoIP environments. The CAPTCHA were classified based on their attributes into four categories: (a) vocabulary, (b) background noise, (c) time, and (d) audio production. Afterwards, the evaluation took place where the CAPTCHAs were utilized on the above mentioned attributes. The evaluation process was based on the fact that CAPTCHAs must be easy for human users to solve, easy for a tester machine to generate and grade, and hard for a software bot to solve. Therefore, the final evaluation was made by two means; namely, by user tests (~60 persons) and by two bots configured to solve audio CAPTCHAs. The evaluation process proved that a) the current CAPTCHA implementations are not adequate, meaning that every implementation is either too easy or too difficult to be solved by both users and bots, and b) the implementation attributes of some CAPTCHAs, like long vocabulary (> 8 characters) and language requirements (native vs. non-native English speakers), affects negatively the users' success rate (~40%) in most cases.

In conclusion, these experimental studies present that solving audio CAPTCHA is particularly problematic. The above results indicate that there is a need and the potential to create more appropriate challenges that will allow for fewer problems in solving the CAPTCHA challenges. In order to achieve more efficient CAPTCHA challenges, one needs to remove the connection between the difficulty of CAPTCHA solving for humans and for bots respectively. This disconnection will allow for the creation of challenges that are more appropriate and consequently easy for humans to solve, without reducing the difficulty for bots.

In this work, we propose adapting the CAPTCHA challenges to the person's characteristics. Utilizing these characteristics we expect to provide challenges that are easier to solve for humans, without aiding the bots at the same time, or even making the challenges harder for bots. The focus of this work is to provide the mechanism that will choose and deploy adaptive audio CAPTCHAs. This presupposes that it is feasible to generate such audio challenges. We consider this task of generating the adaptive audio CAPTCHAs an interesting but also viable requirement, which however is outside of the scope of this work.

Taking into account the person's characteristics during CAPTCHA generation, brings about privacy concerns that need to be addressed. There has been significant progress on the subject of accountable privacy preserving services during the past decade. The privacy-preserving techniques used in the proposed system are closely related to accountable anonymous communication systems (Diaz and Preneel, 2007), anonymous credential systems (Camenisch and Pfitzmann, 2007) and electronic identity cards (Poller et al., 2012, Deswarte and Gambs, 2010). Using cryptographic tools, all these systems aim at providing their functionality while protecting users'

privacy. Similarly, we utilize existing cryptographic primitives to create a privacy-preserving personalized CAPTCHA system. Using certificates, the system allows users to prove attributes about themselves, without revealing their identity.

2 Concepts and Architecture

2.1 Problem Statement and solution overview

The selection of an appropriate CAPTCHA challenge that successfully distinguishes between human users and bots is a challenging task, since generic challenges often pose difficulties to human users as well. Lowering the difficulty level of the challenges consequently allows for bots to solve them as well. Therefore, we believe that a method is needed to tailor CAPTCHA challenges closer to the human user, without lowering the difficulty level for bots. The goal of this work is offering such customized CAPTCHA challenges. Overall, this work does not aim at making CAPTCHA challenges generally easier, rather, it aims at proposing a method to create more appropriate, effective and fair challenges for the users.

The ability for challenge adaptation protects the users from unauthorized use of their accounts (hijacking) and attempts to impersonate them. Additionally, the combination of certificates and the CAPTCHA test, protects the VoIP system from the use of stolen certificates and from users that misuse their certificates for making SPIT calls. Therefore, this work does not aim at making the CAPTCHA test obsolete by using the user certificates. Using the certificates, callers can assert that they have certain characteristics, but this is also verified during the CAPTCHA test.

2.2 Discrimination issues concerning CAPTCHA challenges

Traditionally, problems of accessibility to IT applications and services were addressed by adapting their design to the so-called “average or typical user”, a feature that actually does not exist. CAPTCHAs had been initially recommended and implemented without taking user needs and (dis)abilities as well as accessibility issues into consideration. However usability and accessibility are seriously affected by currently most used visual CAPTCHAs as they pose problems to blind, visually impaired or dyslexic users and, in general, users with disabilities (May, 2005).

Indeed, a CAPTCHA test, which cannot be solved due to the mental or physical disabilities, language, genre, age or even cultural differentiation of the challenged user, interferes with her communication rights (access to and use of IT means) and raises significant discrimination issues (Basso and Bergadano, 2010). A person who cannot respond to a CAPTCHA test on the ground of a disability is discriminated both as subscriber/user of a communication service and as personality, who faces barriers to her communicative interaction with other persons. The use of such a SPIT detection mechanism impairs her right to free communication and consequently the legally embedded right to receive and impart information (Marias et al., 2007).

The United Nations Convention on the Rights of Persons with Disabilities identifies accessibility as one of its general principles and states that States Parties shall take appropriate measures to promote access for persons with disabilities to new information and communications technologies and systems. In many countries including US and the EU countries, legislation in place has to ensure that products and services are accessible and usable by as many users as possible, including people with disabilities and aged persons.

In order to face and/or limit the discriminatory effect of CAPTCHA tests, they should be accessible und usable by all human users, regardless of their cognitive, physical, sensory or cultural characteristics (Fritsch et al., 2010). Apart from having the possibility to switch to a new challenge involving different sensory abilities, the introduction of personalized profiles that take into account the users' diversity, needs and preferences is not only at the core of the inclusive design approach (Fritsch et al., 2010) but seems to be an appropriate response to discrimination concerns.

This approach engages the user in the definition of challenges and tests and considers her needs and abilities. At the very centre of personalized services is the user profile or personal profile, which is a collection of the user preferences and data. However the personalized CAPTCHA service must be designed in a way that allows the user to use and have access to it, while determining when and who should get knowledge about her preferences and /or disability status (Fuglerud et al., 2009). This requirement derives both from the dignity principle and the privacy rights of individuals.

By definition personalized profiles, i.e. in our case personalized CAPTCHAs, require collection and use of personal data (age, education level etc) that may also be sensitive (medical data, disabilities, cultural/religion), affecting the privacy rights of the concerned users. By referring to privacy in this paper we focus on the right of the individual to be in control of the information concerning her so as to formulate conceptions of self, values, preferences, goals and to protect her life choices from public control, social disgrace or objectification.

2.3 PrivCAPTCHA Architecture

In this Section the components of the proposed architecture are described.

2.3.1 Cryptographic Building Blocks

In order to achieve the privacy-preserving attributes of the proposed system (PrivCAPTCHA), we use the cryptographic building blocks presented in the following paragraphs. These building blocks are used as high-level components of our architecture and therefore it is not needed to examine closely the internals of their functions.

2.3.1.1 Anonymous Credentials

Anonymous credentials (Camenisch et al., 2011) allow users to acquire credentials and demonstrate them without revealing their identity. Using the private credential system described in (Camenisch and Pfitzmann, 2007), individuals can use different unlinkable pseudonyms, based on the same credential issued by an identity provider. The private credential system can also provide certified attributes by the identity provider, for the individual to selectively reveal attributes (e.g. their age range, based on their date of birth).

2.3.1.2 Portfolio: a data structure where certified user data is stored.

The user's certificates are stored in a personal portfolio that resides at the owner's side, similar to the one proposed in (Tasidou and Efraimidis, 2012). The contents of a user's portfolio include:

- Certificates of demographic data and personal characteristics, e.g., age, education level, disabilities.

- Certificates of successful CAPTCHA tests issued by the CAPTCHA service. This transaction history can be used to provide further evidence to the CAPTCHA server that the user is human and non-malicious.

2.3.2 Entities

The entities that participate in the proposed system are the following:

The Identity Provider (IDP). Users obtain their credentials by the IDP, by registering an identifier (e.g. their social security number) and a pseudonym P. The IDP is considered a trusted third party (like a passport authority) that retains the user information together with their pseudonym. The IDP does not need to be a single entity, but can be a distributed service, to achieve better service availability and enhanced security.

The User (U). In our system the users are considered the VoIP service users. All can act both as callers and callees. When acting as callers, their portfolio information can be used to receive personalized CAPTCHA challenges as illustrated in Figure 1.

The CAPTCHA server, that also acts as a verifier for the Anonymous Credential System. Moreover, the CAPTCHA server automatically generates the CAPTCHA challenge, evaluates the provided answer and sends back whether the answer is correct or not.

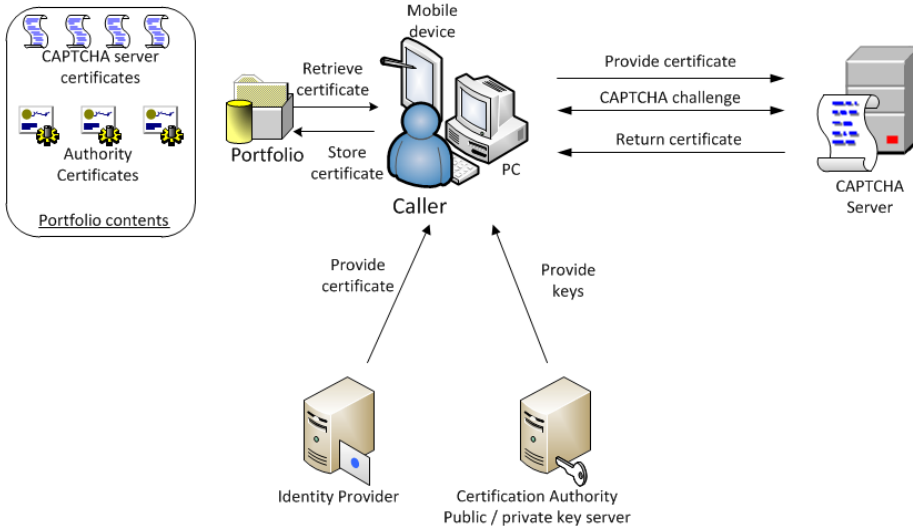

Figure 1: PrivCAPTCHA architecture

The entities of the PrivCaptcha system and their interactions are shown in Figure 1. The user's profile is stored at the personal portfolio residing at the user's side. After registering with an IDP and obtaining her certificates, the caller can prove some attributes to the CAPTCHA server and receive personalized CAPTCHA challenges.

According to the proportionality principle the IDP retains no more data than that strictly required to serve the personalized CAPTCHA service. The IDP entity combines the retained data with a pseudonym in order to protect the identity of the user and it is not allowed to reveal or to use this data for any other purpose , with the exception of law enforcement purposes if and to the extent that is provided by the respective law. The proposed architecture provides personalized and effective VoIP CAPTCHAs while preserving the privacy and communication rights of the user.

2.4 System Functionality

In this section we will describe the general functionalities of the system. Due to lack of space we will not describe the use of the cryptographic tools in detail. Besides, we adhere to the descriptions of the primitives as proposed by their authors.

The main functions of the system are the following:

1. Providing certificates and proving attributes to the CAPTCHA system. After acquiring her credentials by the IDP the user (U) can begin using them to prove attributes to CAPTCHA services in order to receive personalized challenges. U needs to prove to the CAPTCHA service that she has a valid credential C, which verifies that U has a certain attribute.

The verification mechanism is based on efficient zero knowledge proofs, like the ones used in (Camenisch and Pfitzmann, 2007).

2. CAPTCHA generation and outcome. After receiving and verifying the user's certificate, the CAPTCHA service generates the appropriate CAPTCHA challenge, according to the user characteristics. The main two characteristics that the CAPTCHA server takes into account are:

   Language requirements. Based on the users' native language, the CAPTCHA server can provide the appropriate challenge. A certificate on the user's ethnicity can be used to determine the language parameter. As the language itself is not as sensitive as the ethnicity information, language selection can be supported without the need for a certificate, allowing users to select languages other than their native. We believe that the language choice will not diminish the system's effectiveness regarding false positives (i.e., bots will not be able to solve the tests more easily).

   Age. The age of the user affects his ability to answer correctly difficult challenges. If the user is too young or too old, then the CAPTCHA challenge should be adapted to contain less characters or words to be recognized.

      Additional user characteristics that can be considered are mental and physical disabilities (e,g, mobility issues, visual impairment), education level (e.g, literacy) and learning disabilities (e.g. dyslexia).

3. CAPTCHA server certificates. Upon successful completion of a CAPTCHA challenge, the CAPTCHA server sends the user a certificate, attesting that this user did make a legitimate communication. This certificate is inserted into the portfolio and can be used later from U to prove that she is a previous legitimate user to the CAPTCHA service.

The above functionalities are prone to misuse and malicious behavior on the part of the user. In the following Section we address the main issues that have been identified for the system.

2.5 Incident response requirements

Designing a system on a user-centric driven security basis requires that the system is robust in a sense that the user is not significantly exposed during a security failure. In the proposed system we have identified the following aspects and requirements for incident response and escalation procedures in an event of a security failure.

Tolerance to false positives. We can in principle consider that false positives carry a minor security impact. The event of the case of a bot answering successfully the audio CAPTCHA challenge will be detected by the destination/callee and the service should maintain the facility for the callee to report/redirect the call for further

logging and analysis. Responding to false positives is a good example of active user participation in the security process. Regarding CAPTCHA server certificates, in case of false positives, a revocation procedure can be followed upon receipt of the callee report.

Tolerance to false negatives. Rejecting a legitimate call request after a failed audio CAPTCHA attempt is an event of major significance. Therefore the underlying security parameters are expected to be set on a level where the false negatives are minimised despite the drop in security. Besides, giving priority to user acceptance over security is part of the user-centric system design practice. In addition, there needs to be a continuous evaluation similar to vulnerability assessment practices. More specifically, as a security administrator must be informed and proactively search for new vulnerabilities affecting the system she is responsible to manage, the audio CAPTCHA engineer must keep the system up to date with the state of the art research in order to maintain the optimum level of security versus user acceptance.

Tolerance to CAPTCHA server certificate misuse. The certificates provided by the CAPTCHA server can be (un)intentionally misused by users to exploit the system. In case of reported malicious use of these certificates, revocation methods (Camenisch et al., 2011) can be examined.

Correlate system failures with SPIT results. A threat management system should be implemented and the audio CAPTCHA service should be placed in the wider system security context in order to identify threat vectors that may target the CAPTCHA but also to exploit the system as a whole.

Reputation management. Reputation mechanisms introduce a number of security issues and should these become part of the audio CAPTCHA service, reputation misuse should be addressed with well defined escalation procedures. The proposed system can adopt published procedures and controls for reputation management.

3 Discussion and Conclusions

In this work, we propose a user-centric, privacy-preserving VoIP CAPTCHA adaptation method. The PrivCAPTCHA architecture combines existing cryptographic technologies, which provide strong privacy guarantees, utilized under a new context. The proposed system aims at providing an improved CAPTCHA service that is more appropriate for and fair to the human users and overall more effective. Although high-level descriptions of the system functionalities are provided in this work, it would be interesting to implement them within the VoIP protocol. The introduction of cryptographic tools is expected to introduce a computational overhead into the audio CAPTCHA application. We expect this overhead to be tolerable for modern computational platforms possibly combined with appropriate performance optimization techniques. Moreover, although we mainly consider CAPTCHA challenges for VoIP calls in this work, we believe that this idea can be useful for providing a general mechanism for CAPTCHA adaptation according to the users' characteristics.

4 Acknowledgements

This work was performed in the framework of and funded by the GSRT/CO-OPERATION/SPHINX Project (09SYN-72-419) (http://sphinx.vtrip.net). A. Tasidou P. S. Efraimidis and V. Katos are partially supported by national (ETAA) funds.

5 References

Ahn, L. v., Blum, M. and Langford, J. (2004), "Telling humans and computers apart automatically", Communications of the ACM, Volume 47, Number 2, pp. 56-60, ISSN: 0001-0782.

Basso, A. and Bergadano, F. (2010), "Anti-bot Strategies Based on Human Interactive Proofs", in Stavroulakis, P. and Stamp, M. (Eds.) Handbook of Information and Communication Security, Springer, Berlin / Heidelberg, ISBN: 978-3-642-04117-4.

Bigham, J. P. and Cavender, A. C. (2009), "Evaluating existing audio CAPTCHAs and an interface optimized for non-visual use", Proceedings of the 27th international conference on Human factors in computing systems, Boston, MA, USA, 2009, pp. 1829-1838.

Bursztein, E., Bethard, S., Fabry, C., Mitchell, J. C. and Jurafsky, D. (2010), "How good are humans at solving CAPTCHAs? a large scale evaluation", Proceedings of the 2010 IEEE Symposium on Security and Privacy, Oakland, California, USA, 2010, pp. 399-413.

Camenisch, J., Dubovitskaya, M., Kohlweiss, M., Lapon, J. and Neven, G. (2011), "Cryptographic Mechanisms for Privacy", in Camenisch, J., Fischer-Hübner, S. and Rannenberg, K. (Eds.) Privacy and Identity Management for Life, Springer, Berlin / Heidelberg, ISBN: 978-3-642-20317-6.

Camenisch, J. and Pfitzmann, B. (2007), "Federated Identity Management", in Petković, M. and Jonker, W. (Eds.) Security, Privacy, and Trust in Modern Data Management, Springer, Berlin / Heidelberg, ISBN: 978-3-540-69861-6.

Deswarte, Y. and Gambs, S. (2010), "A Proposal for a Privacy-preserving National Identity Card", Transactions on Data Privacy, Volume 3, Number 3, pp. 253-276, ISSN: 1888-5063.

Diaz, C. and Preneel, B. (2007), "Accountable Anonymous Communication", in Petković, M. and Jonker, W. (Eds.) Security, Privacy, and Trust in Modern Data Management, Springer, Berlin / Heidelberg, ISBN: 978-3-540-69861-6.

El Sawda, S. and Urien, P. (2006), "SIP Security Attacks and Solutions: A state-of-the-art review", Proceedings of the 2nd International Conference on Information & Communication Technologies: From Theory to Applications, Damascus, Syria, 2006, pp. 3187-3191.

Fritsch, L., Fuglerud, K. and Solheim, I. (2010), "Towards inclusive identity management", Identity in the Information Society, Volume 3, Number 3, pp. 515-538, ISSN: 1876-0678.

Fuglerud, K., Reinertsen, A., Fritsch, L. and Dale, Ø. (2009), "Universal design of IT-based solutions for registration and authentication", Tech. report: DART/02/09, Norwegian Computing Center, Oslo, 2009.

Hernández-Castro, C. J., Ribagorda, A. and Hernández-Castro, J. C. (2011), "On the Strength of Egglue and Other Logic CAPTCHAs", Proceedings of the International Conference on Security and Cryptography (SECRYPT), Seville, Spain, 2011, pp. 157-167.

Marias, G. F., Dritsas, S., Theoharidou, M., Mallios, J. and Gritzalis, D. (2007), "SIP Vulnerabilities and Anti-SPIT Mechanisms Assessment", Proceedings of 16th International Conference on Computer Communications and Networks, Honolulu, Hawaii, USA, 2007, pp. 597-604.

May, M. (2005), "Inaccessibility of CAPTCHA. Alternatives to visual Turing tests on the Web.", http://www.w3.org/TR/turingtest/, (Accessed November 2005).

Poller, A., Waldmann, U., Vowe, S. and Turpe, S. (2012), "Electronic Identity Cards for User Authentication; Promise and Practice", IEEE Security & Privacy, Volume 10, Number 1, pp. 46-54, ISSN: 1540-7993.

Soupionis, Y. and Gritzalis, D. (2010), "Audio CAPTCHA: Existing solutions assessment and a new implementation for VoIP telephony", Computers & Security, Volume 29, Number 5, pp. 603-618, ISSN: 0167-4048.

Tasidou, A. and Efraimidis, P. S. (2012), "Using Personal Portfolios to Manage Customer Data", in Garcia-Alfaro, J., Navarro-Arribas, G., Cuppens-Boulahia, N. and De Capitani di Vimercati, S. (Eds.) Data Privacy Management and Autonomous Spontaneous Security, Springer, Berlin / Heidelberg, ISBN: 978-3-642-28878-4.

Walsh, T. J. and Kuhn, D. R. (2005), "Challenges in securing voice over IP", IEEE Security & Privacy, Volume 3, Number 3, pp. 44-49, ISSN: 1540-7993.