Evaluating the privacy properties of secure VoIP metadata

João S. Resende[0000−0003−0125−4240], Patrícia R. Sousa[0000−0002−0268−9134], and Luís Antunes[0000−0002−9988−594X]

Department of Computer Science, Porto, Portugal
{jresende,psousa,lfa}@dcc.fc.up.pt

Abstract. Some governments do not consider metadata as personal data, and so not in the scope of privacy regulations. However, often, metadata gives more relevant information than the actual content itself. Metadata can be very useful to identify, locate, understand and manage personal data, i.e., information that is eminently private in nature and under most privacy regulation should be anonymized or deleted if users have not give their consent. In voice calls, we are facing a critical situation in terms of privacy, as metadata can identify who calls to whom and the duration of the call, for example. In this work, we investigate privacy properties of voice calls metadata, in particular when using secure VoIP, giving evidence of the ability to extract sensitive information from its ("secure") metadata. We find that ZRTP metadata is freely available to any client on the network, and that users can be re-identified by any user with access to the network. Also, we propose a solution for this problem, suitable for all the ZRTP-based implementations.

Keywords: Metadata · VoIP · Privacy

1 Introduction

Nowadays, voice calls are so common that we carry on our pockets a device that we use mainly for this purpose. In the recent years, it has been disclosed that some countries can massively tap this calls, staging the pave for secure voice calls. However, often despite the privacy of the "content" of the data "being preserved", the metadata has serious implications on privacy. Metadata leakage is not about the information content of published data but the properties of it that makes disclosures of sensitive personal information [7]. The properties can be size or location (in the case of an image, for example, modern smartphones (and many digital cameras) embed GPS coordinates in the photos) [8]. The user's privacy can be compromised through metadata as the information can identify the user (directly or indirectly). NSA General Counsel Stewart Baker has said, "*metadata absolutely tells you everything about somebodys life. If*

you have enough metadata, you dont really need content." [10]. In fact, even the indirect information, the interconnection between metadata can provide a complex profile of the person in question.

A recent study by *A. Gruber* [1] presents an application that characterizes the behavioral patterns of suspect users versus non-suspect users based on metadata usage such as call duration, call distribution, interaction time preferences and text-to-call ratios, while avoiding any access to the content of calls or messages. This was based on the traditional communications and the metadata provided by the smartphone.

However, with the evolution of the communications industry, many alternatives have been offered to the business market. Among these options are the telephone exchanges, corporate cell phones, conventional telephones and Internet Protocol (IP). The options offered by telephones, came to simplify and streamline organizational communications. The costs of communication between people from different countries by telephone connection are very high, in addition to the cost with the subscription and the costs with the additional minutes. Not only for these reasons, but in this sense, the technologies currently available can help save money and Voice over Internet Protocol (VoIP) technology is becoming more popular. VoIP transforms analog audio signals into digital data that can be transferred over the internet. With the spread of the internet, this technology has become increasingly common and we can easily see it in tools like Whatsapp, Skype, Facebook Messenger, among others.

The goal of this work, is to analyze a sub-field of metadata based on a protocol used in some well known secure VoIP applications (such as Linphone [3], Silent Phone [4] [5]) called ZRTP [6] ("Z" is a reference to its inventor, Zimmermann; "RTP" stands for Real-time Transport Protocol).

This paper is organized as follows: Section 2 describes the related work with some attacks already discovered of ZRTP and some literature about the metadata information extraction and how privacy is concerned with this type of information leak. Section 3 describes the problem that we are addressing. Section 4 state the implementation, setup and demonstration of the information extracting by metadata and ZRTP identifier (ZID). In section 5, we present a possible solution for the problem of the information leakage. Finally, the last section presents the conclusions of the work and some future work in this area.

2 State of the Art

This section provides an overview of the literature, focusing on ZRTP attacks and metadata analysis in general, with special focus on private information leakage.

ZRTP is a end-to-end secure communication protocol that contains a session set-up phase used to agree on key exchange and parameters for establishing Secure Real-time Transport Protocol (SRTP) sessions. The Diffie-Hellman (DH) method is a specific cryptographic algorithm for key exchange based on discrete logarithms. These keys contribute to the generation of the session key and parameters for the SRTP sessions. Although ZRTP initially needs to use a signalling protocol, such as a Session Initiation Protocol (SIP)[1], the key negotiation is performed only by ZRTP. The DH algorithm, alone, does not provide protection against Man-in-the-Middle (MitM) attacks. In order to authenticate both peers in the key exchange on ZRTP, a Short Authentication Strings (SAS) is used that is distributed to both phones and compared verbally by both ends. If the SAS is the same, both users must press a button in order to accept the key. The SAS is destroyed in the end of the call, but if the users don't verbally compare the SAS, after the first call the media stills with authentication against a MiM attack based on a form of key continuity, if both ends have in a previously call accepted the SAS, so after the first call the users don't need to continuously check the SAS.

The ZRTP provides the user two important features to become more efficient and robust: Preshared mode and key continuity. In the Preshared mode, both party can skip the DH calculation if they have established a ZRTP media session before, and both users have verified if the SAS displayed in the previous call matched in both ends. Preshared mode uses the previous shared secret, to establish the next call and holds forward secrecy. Forward secrecy assures that, if one attacker gains access to the keys present in one device, he is not able to decrypt previous communications, as the shared secret is replaced in each new call from the same pair of users. So, if the attacker gains access to the cached secret, he can gains access to all future communication while the Preshared mode remain to be used if he can intercepts all the call from that moment on. The key continuity features, caches some hashed key information to be used in the future call and, exchange with each session. If Mallory is capable of steal shared secrets caches from one user, the user just have one opportunity

[1] SIP is a third party server that allows the peer discovery and negotiation, in the case of ZRTP does not interact with the key negotiation

to perform a MiM/eavesdropping attack in the next session. If Mallory misses the interception the shared secret is updated and the opportunity to intercept all the calls from their on is lost.

As the ZRTP is used in many secure call applications, it is critical to know if there is a possibility of perceiving which calls are being made by the user, who the calls are to be made to, call duration, etc. Often this subject is addressed only at the content level of the call, but sometimes realizing these call patterns between users is also critical to the privacy of each of them. This analysis will allow to see if it is possible to re-identify people and their communications.

2.1 Attacks on ZRTP

As presented in [16], it is possible to perform a MiTM attack on ZRTP. Instead of cracking the DH key exchange, Mallory - the malicious attacker, tries to force Alice and Bob (the honest participants) connect directly to her, without knowledge. This way, Alice and Bob have two different connections with Mallory and she just needs to relay the packets from one connection to the other. There are different session keys and different SAS in these two connections, meaning that if Alice or Bob decide to exchange the SAS, Mallory will be detected (because the SAS will be different in both devices). To solve this, the same document [16] suggest that Mallory can avoid this detection by Direct relay + Imitation SAS, Direct relay + Slide-stepping SAS and Masquerade. To be able to perform these attacks, Mallory always needs to perform a new DH key exchange and for that, the authors use different ZID to make the end users exchange new keys. The authors also suggest that a different ZID in a new call is normal, it can be originated by a new installation on the device or by one exchange of device, that always will end with a new ZID.

The most recent attack on ZRTP presented by *Schürmann et al.* [15] is focused on both implementation and theoretical analyses by applying practical usage to applications such as *Linphone* and *CSipSimple*. They present some security breaches starting by a *CVE-2016-6271* that overcome the security limitations imposed by the ZRTP protocol because it does not implement the packet verification of some parameters. However, the paper also addresses the protocol structure from the ZRTP where the authors explore a vulnerability from the ZID. In this vulnerability, it is possible explore a MiTM after the first call, so, an attacker just needs to perform a call for the victim beforehand. To solve this issue, the authors use the SIP URI along with the ZID of the client to lock that attack. *Dim-*

itrios Alvanos et al. [24] also present a document with analysis on security features of ZRTP VoIP clients, more specifically, *Linphone* client.

2.2 Metadata Information Extraction

There are several studies evaluating the privacy impact of metadata information extraction. Smart services and experiences are based on high-dimensional metadata from users. For example, in Netflix or Amazon, metadata is used by commercial algorithms to help users become more connected, productive, and entertained [2].

There exist several algorithms to re-identifying people based on metadata, namely based on human behavior, such as the credit-card metadata [12], *Like credit card and mobile phone metadata, Web browsing or transportation data sets are generated as side effects of human interaction with technology, are subjected to the same idiosyncrasies of human behavior, and are also sparse and high-dimensional (for example, in the number of Web sites one can visit or the number of possible entry-exit combinations of metro stations).* In another example, *Arvind Narayanan et al.* [13] apply de-anonymization algorithms to a Netflix Prize database that anonymizes the movie ratings of 500,000 Netflix subscribers. They have demonstrated that they can re-identify the subscribers only knowing little about each subscriber.

These are some examples of re-identification based on human behavior. This motivates our work, as we are interested, in some way, with human behaviours. To the best of our knowledge, the metadata of VoIP was never studied, and by nature it may be quite sensitive if leaking information about the calls and participants.

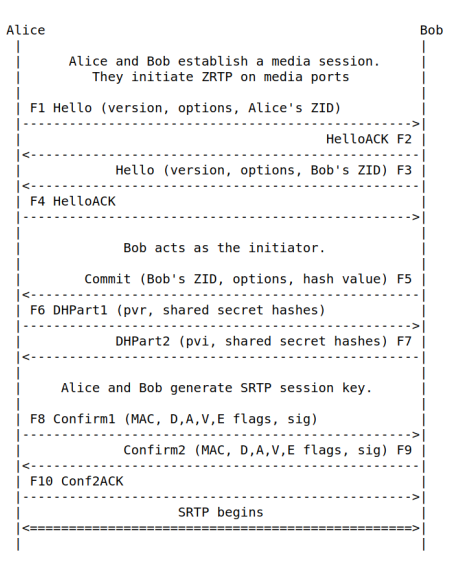
Still, there are some work done on telephone metadata [19–23]. *Jonathan Mayer et al.* [9] study the privacy properties of telephone metadata to assess the impact of policies that distinguish between content and metadata. They claim that there are significant privacy flaws associated with telephone metadata because it *is densely interconnected, easily re-identifiable, and trivially gives rise to location, relationship, and sensitive inferences. Marco Furini* [11] also states that *the high availability of geolocation technologies is changing the social media mobile scenario and is exposing users to privacy risks.. Alexandre Pujol et al.* analyze the metadata of instant messaging services in order to retrieve sensitive knowledge. The work assumes that the attacker has a full access to the server, in order to retrieve metadata. Also, this work proposes a solution for this metadata leak based on Oblivious RAM model.

3 Problem statement

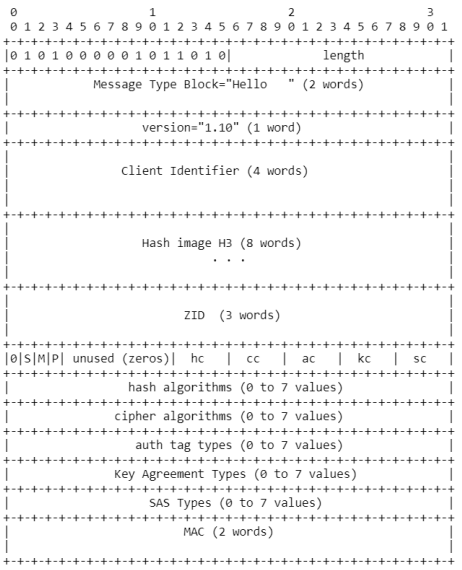
The privacy risk, as we have already mentioned, is not only concerned with the ciphered content of the voice call, but rather with the care taken with the metadata. Often, metadata reveals more information than the content itself (about the name, location, etc.), these are the data that lead us to profile the participant and to re-identify him.

VoIP calls have not yet been analyzed at the metadata level, and so, we decided to analyze the ZRTP, where calls are encrypted in both ends, however, does the metadata give us some sensitive information or is is possible to infer information that should be anonymized?

ZRTP is a cryptographic key-agreement protocol meant to negotiate the keys for encryption between two end points in VoIP telephony. This key agreement (Figure 1a) can be divided into four steps: Discovery, Hash Commitment, DH exchange and Key Derivation and Confirmation.



(a) ZRTP Key Agreement Packet Flow [6]



(b) "Hello" Message Packet Format [6]

Fig. 1: ZRTP

These phases are described in the RFC of ZRTP [6], and we focus on the discovery phase, that exchange some information that is sensible. In this phase, the initiator and the responder exchange their ZRTP identifiers, as well as information about each others supported ZRTP versions,

hash functions, ciphers, authorization tag lengths, key agreement types, and SAS algorithms through the packet Hello (figure 1b). One important field is the unique 96-bit random ZID that is generated once at installation time. This field is important because it is used to index the cryptographic materials used by the key continuity and forward secret properties of the protocol in each device.

By having an identifier such as the ZID, it is possible to perceive call patterns with network sniffing, as the ZID is unique and is shared with both the callers (receiver and sender ZIDs). In addition, the ZID can be combined with metadata (such as the duration of the call) to know how long the call between the parties lasted. So far, even though we infer from the ZID which calls exist and between whom, as well as the duration of the call, we could not connect the ZID to the physical person. Note that, the ZRTP packets are used without extra encryption, meaning that a content of a packet *Hello* can be seen by any person while the transfer is being processed from the origin to the destination through the server. The reason is that ZRTP is a protocol design for end-to-end communication, so, this should not exchange crypto material by a third party server unless needed. This attack can be conducted by an attacker on the same network of the victim, by having specially authorizations to access to the network, by an ISP or governance agencies.

4 Extracting the information

This section focuses on the study of how this information can increase the risk of re-identification or profiling users, creating a traceability problem regarding the privacy of the users.

4.1 Implementation

To test VoIP communications based on ZRTP, we use the *Linphone* [3] client (available on the Google Play).

To simulate the real environment, we start by using a public SIP server (*sip.linphone.org*). However, during the tests, we detect that all the communications are made with the server and not peer-to-peer. The figure 2 represents a normal call between Alice and Bob where the packets are sent/receive by the server that brings an extra problem regarding security, because users communicate always with the server, so the eavesdropping process can be taken not only on the local network (similar to our setup) but also during the handling of the packet up to and from the server.

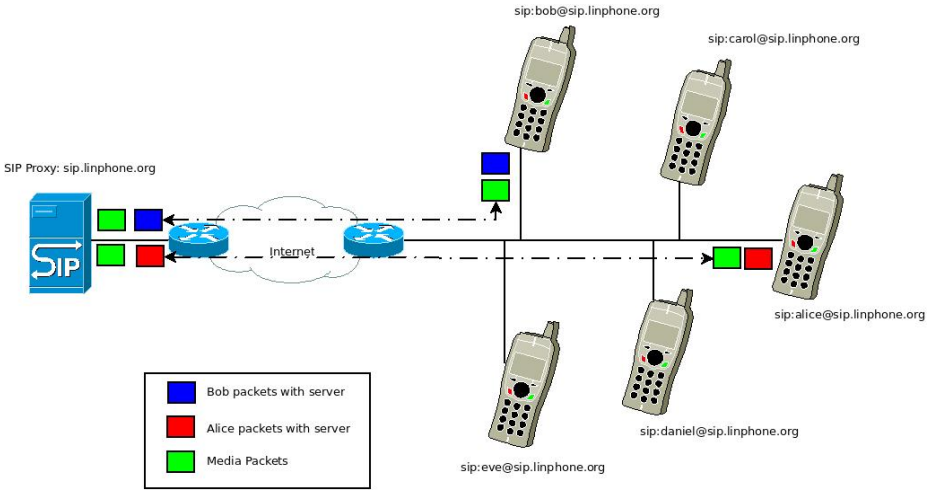


Fig. 2: ZRTP peer-to-peer communication

In order to create a communication pattern dataset, we first started by implementing a sniffer. We use the *tcpdump* to collect the information from all the network packets that contain all the communications between the users and the SIP server to establish a peer-to-peer connection. Then, we use the *scapy* tool [14] to process the information (packets) stored by *tcpdump* and detect the ZID's of a given connection between two users. The creation of the dataset was based on the following steps:

1. Using *scapy*, we receive and manipulate the packets from *tcpdump* starting by search for the field "*Message Type Block*" to see if we got the packet "Hello" (figure 1b);
2. Then, with the tool, we select the field ZID of both ends of the communication and store it in the dataset;
3. During the transmission of the packets, we also use *scapy* to store the initial and final time of the transmission, in order to get the duration of the call;
4. Finally, we store in the dataset a field of each communication with this format: <Alice's ZID, Bob's ZID, Initial Call Time, Call Duration Time>.

4.2 Setup

We deployed a laboratory environment simulating a company with five employees, where one of them just arrived to the company. The employees

are represented by five mobile devices (with Android operating systems from version 19 to 27) connected to a wireless device. Also, an external sniffing network device (represented by the *tcpdump* that just dumped the information to a *Wireshark* type file) is on the network.

4.3 Demonstration

For demonstration purposes we assume the following scenario: In a company there are four employees and a new one arriving to the company. The new employee wants to know more than he should about communications within the company (who communicates with whom, and for how long each call was made). Figure 3 illustrates this scenario. In this case, each phone has the ZIDs of the caller.

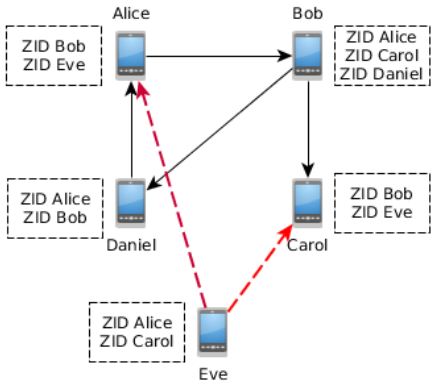


Fig. 3: ZRTP Communication

When Eve arrives at the company, despite seeing the communications and the ZID of each one with a network sniffer, it can not identify each user. However, assuming that Eve calls Alice and Carol because once she gets each of the ZIDs from these two people, she automatically can identify the history of these two users. The history contains the number of calls and the duration of the calls between these two users or between these users and the remaining. From there, the privacy of Alice and Carol are compromised, as Eve gets to know too much about the communications that each of them perform.

With this example of Eve, we can assume that if the users in the network (all employees) want to do network sniffing, they will get the

same information about other employees. Therefore, the privacy of these users is compromised. In addition, for the purposes of forensics, we know that it is enough to know that two persons have communicated with each other and their respective duration of the call. Therefore, it is critical to have ZID privacy so that there is no such information leakage, as well as the metadata that is generated over the duration of the call.

Finally, we can also identify users by the behaviour in the real life. This way, Eve can get the match of the ZID with the physical person by searching patterns for example, isolating by the employee in vacations or time of the day.

5 Proposed Solution

To be able to solve the information leakage, we can follow two possible solutions. Both need to use the SIP URI, that is a string used to identify the user towards a SIP server (an example can be *sip:alice@sip.linphone.org*). In any possible solutions, this SIP URI is needed to fix the problem as stated by *Schürmann et al.* [15].

5.1 Scenario 1

We create the first scenario to solve the privacy issue in all the calls after the first one with each peer. To encrypt the ZID, we propose the use of the derivation of the cryptography material stored on the local phone. However, in the first call between two peers, we do not have the cryptographic material. For this reason, the ZID is not encrypted in the first call.

With this solution, the attacker can only trace the first call between each communication between two users. This protects the privacy (after the first call) of users by mitigating the problem of leakage of the duration of the calls and also the number of calls between each pair of users, improving the protection against traceability.

5.2 Scenario 2

The second scenario is focused essentially on the use of SIP URI. As we already describe in the Related Work Section, there are a vulnerability related to MiTM where the proposed solution is the use of SIP [15]. In order to secure all the protocol, and as we have to use SIP to be secure, we can use it to discover the peers instead of use ZID. So, SIP URI is

used to look up retained shared secrets from previous ZRTP sessions with the endpoint, to replace the functionality of ZID. In brief, it is used as an index into a cache of shared secrets that were previously negotiated and retained between the two parties.

Note that, we use SIP URI only for identification of peers, that is, we are only dependent of a third party in the discover phase.

5.3 Implementation

Based on the two proposed scenarios, we decide to follow Scenario 2 and implement it. This decision is mainly focused on the mitigation of the metadata leakage problem even at the first call.

In order to test the proposed solution, we implement a prototype-/patch based on *ZRTPCPP* [17] library with *PJSIP* [18]. However, we must ensure that the users of *ZRTPCPP* can communicate with the users of *Linphone* for example, or any other application that supports ZRTP. As a requirement we consider also that the implementation must be compliant with the RFC of the ZRTP [6], in order to be suitable to all the implementations of the protocol. For this reason, it is important to follow the protocol and the packets format presented in the RFC, making the implementation interoperable. Also, the implementation must be secure and privacy by design, where we block the traceability issue and not leak any other information maintaining the communications secure and private.

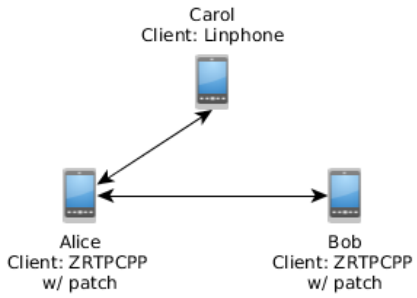


Fig. 4: Different ZRTP Client Communications

Our patch propose the replacement of ZID by the SIP URI. In a normal call between two users using our scenario (Figure 4) , it will be granted

that both clients will store the SIP URI from the SIP layer in order to store the credentials values and maintain the preshared mode. One of the most important features of preshared mode is the key continuity, that is important because the validation of the SAS is not necessary after the first call. However, if we have a call between two different clients (Alice and Carol - figure 4), the ZID field is required by Carol (she is not using our scenario). For this reason, we keep this field with a random value to be compliant with the RFC of ZRTP. In the communication between Alice and Carol, the users do not have the key continuity/preshared mode properties, so, it will be used the DH key agreement in all calls. This means that it is required the validation of the SAS in all the communications, even after the first call (it looses the key continuity because the ZID is random, and therefore, the ZID is different in all the calls).

The traceability problem is mitigated with our solution because what the attacker sees on the network is that Alice is always calling a "different" user in each call to Bob. So, the attacker does not know that Alice is calling Bob and so, it is impossible to trace their behaviour. Regarding Carol, the problem of traceability is still there but mitigated because she is speaking with a patched client that will always produce a different ZID value, so, it is impossible to see who is she calling to.

6 Conclusion

In this paper, we described a metadata vulnerability in secure VoIP calls. We stress that this finding is of utmost importance as these users are looking not only for secure communications but also for privacy guarantees. To make our findings clear we developed a laboratory demonstration of how to explore this vulnerability, along with a state of the art regarding similar problems both in terms of metadata in voice communication and others. Given the high privacy risk, we proposed two solutions. Finally we implement also a VoIP client that implements this security properties based on ZRTPCPP.

The insights and results gained throughout this work highlights the necessity of using open-source resources where researchers can study and deploy this type of approach, in an effort to build a secure open-source ecosystem. The future work will focus in the analyzes of other VoIP protocols and in the exploration of this metadata in a real environment, to understand communications patterns.

Acknowledgements

This work is partially funded by the ERDF through the COMPETE 2020 Programme within project POCI-01-0145-FEDER-006961, and by National Funds through the FCT as part of project UID/EEA/50014/2013.

The work of João S. Resende was supported by a scholarship from the Fundação para a Ciência e Tecnologia (FCT), Portugal (scholarship number PD/BD/128149/2016).

The work of Patrícia R. Sousa and Luís Antunes was supported by Project "NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics/NORTE-01-0145-FEDER-000016", financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

1. Gruber, A., and I. Ben-Gal. "Using targeted Bayesian network learning for suspect identification in communication networks." International Journal of Information Security 17.2 (2018): 169-181.
2. de Montjoye, Yves-Alexandre, et al. "openpds: Protecting the privacy of metadata through safeanswers." PloS one 9.7 (2014): e98790.
3. LinPhone Open source VOIP project (2017) http://www.linphone.org/ [Online; Accessed 29-03-2018]
4. Moscaritolo, Vinnie, Gary Belvin, and Phil Zimmermann. "Silent circle instant messaging protocol protocol specification." Online, White Paper (2012).
5. Silent Circle (2018) https://www.silentcircle.com/ [Online; Accessed 29-03-2018]
6. Zimmermann, Phil, Alan Johnston, and Jon Callas. ZRTP: Media path key agreement for unicast secure RTP. No. RFC 6189. 2011.
7. Greschbach, Benjamin, Gunnar Kreitz, and Sonja Buchegger. "The devil is in the metadataNew privacy challenges in Decentralised Online Social Networks." Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on. IEEE, 2012.
8. Tesic, Jelena. "Metadata practices for consumer photos." IEEE MultiMedia 12.3 (2005): 86-92.
9. Mayer, Jonathan, Patrick Mutchler, and John C. Mitchell. "Evaluating the privacy properties of telephone metadata." Proceedings of the National Academy of Sciences 113.20 (2016): 5536-5541.
10. Cole, David. "We kill people based on metadata." The New York Review of Books 10 (2014): 2014.
11. Furini, Marco, and Valentina Tamanini. "Location privacy and public metadata in social media platforms: attitudes, behaviors and opinions." Multimedia Tools and Applications 74.21 (2015): 9795-9825.
12. De Montjoye, Yves-Alexandre, Laura Radaelli, and Vivek Kumar Singh. "Unique in the shopping mall: On the reidentifiability of credit card metadata." Science 347.6221 (2015): 536-539.

13. Narayanan, Arvind, and Vitaly Shmatikov. "Robust de-anonymization of large sparse datasets." Security and Privacy, 2008. SP 2008. IEEE Symposium on. IEEE, 2008.
14. Scapy: the Python-based interactive packet manipulation program & library, 2015. https://github.com/secdev/scapy/
15. Schrmann, Dominik, et al. "Wiretapping End-to-End Encrypted VoIP Calls: Real-World Attacks on ZRTP." Proceedings on Privacy Enhancing Technologies 2017.3 (2017): 4-20.
16. Petraschek, Martin, et al. "Security and Usability Aspects of Man-in-the-Middle Attacks on ZRTP." J. UCS 14.5 (2008): 673-692.
17. Werner Dittmann, ZRTPCPP, 2018 https://github.com/wernerd/ZRTPCPP
18. PJSIP version, teluu http://www.pjsip.org/
19. Toole, Jameson L., et al. "Tracking employment shocks using mobile phone data." Journal of The Royal Society Interface 12.107 (2015): 20150185.
20. Arai, Ayumi, et al. "Understanding user attributes from calling behavior: exploring call detail records through field observations." Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia. ACM, 2014.
21. de Montjoye, Yves-Alexandre, et al. "Predicting personality using novel mobile phone-based metrics." International conference on social computing, behavioral-cultural modeling, and prediction. Springer, Berlin, Heidelberg, 2013.
22. Chittaranjan, Gokul, Jan Blom, and Daniel Gatica-Perez. "Mining large-scale smartphone data for personality studies." Personal and Ubiquitous Computing 17.3 (2013): 433-450.
23. Zhong, Erheng, et al. "User demographics prediction based on mobile data." Pervasive and mobile computing 9.6 (2013): 823-837.
24. Alvanos, Dimitrios, Konstantinos Limniotis, and Stavros Stavrou. "On the Cryptographic Features of a VoIP Service." Cryptography 2.1 (2018): 3.