Identity Linking in Computer Networks

Libor Polčák[1][a], Ondřej Ryšavý[1][b] and Petr Matoušek[1][c]
[1]*Brno University of Technology, Faculty of Information Technology, Centre of Excellence IT4Innovations, Božetěchova 2, 612 66 Brno, Czech Republic*
{*ipolcak,rysavy,matousp*}*@fit.vutbr.cz,*

Keywords: Cross-Layer Identity Linking, Identifiers, Graph of Identifiers, Applications of Labeled Property Graphs.

Abstract: Lawful interception, network forensics, and security incident investigations require cross-layer linking of identification information to link different activities of a particular person. This paper presents a model called graphs of identifiers that allows cross-layer linking of identifiers detected by various methods. Graphs of identifiers provide operations that link identifiers according to the constraints provided in the queries. The goal is to employ the linking during early stages of the network forensic investigations when an investigator searches for leads. The tools that implement the proposed model are publicly available.

1 INTRODUCTION

Identifiers appearing in the network identify different subjects (Pfitzmann and Hansen, 2010), for example, human beings, legal persons, or computers. For many network-related tasks spanning lawful interception, network forensics, and security incident investigation, it is necessary to link identities of different subjects. The linking is necessary to answer questions such as *Who was using the computer with a particular IP address last Friday at 5 PM?*

An IP address identifies a computer interface. However, an IP address is not stable (Polčák et al., 2014b) and can be hidden behind a network address translation (NAT). At the application layer, users authenticate to various services. They open sessions, each of which is carried over one or more TCP or UDP connections. All communication can be identified with identifiers occurring inside the traffic flow (FIDIS project, 2008b).

This paper proposes graphs of identifiers based on undirected *labeled property graphs* (Robinson et al., 2015). A graph of identifiers can link identifiers used by a particular subject and link different subjects based on the relationships between the subjects. The model and operations in the model described in this paper are available in a tool called *linking*[1].

This work enhances the graph model established by (Polčák et al., 2014b). Comparably, this paper represents identifiers as nodes in a graph. The novelty of the model presented in this paper lays in assignments of key-value pairs to both nodes and relationships. This improves the information that can be revealed by the model:

- Time is an inherent component of graphs of identifiers. Hence, it is possible to track and investigate time-related identifier linking which is crucial for fields such as network forensics and security incident investigations.
- The model supports probabilistic identification methods (Pfitzmann and Hansen, 2010) — methods that are not able to indisputably reveal the subject represented by an identifier. Instead, such method reveals the identity with some degree of certainty. Graphs of identifiers support linking of identifiers detected by probabilistic identification. Hence, during network forensics, it is possible to treat all identification information as inaccurate, see (Casey and Jaquet-Chiffelle, 2017) for more information on the need for careful examination during network forensics.
- The model considers resources such as web pages or chat rooms. The underlying labeled property graphs allow a definition of additional categories of identifiers.
- We provide more operations compared to the original model with the possibility to define even more operations. The operations are defined as walks in the labeled property graphs.

[a] https://orcid.org/0000-0001-9177-3073
[b] https://orcid.org/ 0000-0001-9652-6418
[c] https://orcid.org/0000-0003-4589-2041
[1] https://github.com/polcak/linking

This paper primarily uses the terminology established by (Pfitzmann and Hansen, 2010). An identifier is a unique attribute value of the subject. FIDIS project introduced a concept of natural persons and virtual persons (FIDIS project, 2008a). According to the terminology established by FIDIS, the model established in this paper links identifiers in the virtual world only.

This paper is organized as follows. Section 2 overviews scenarios that benefit from identifier linking. Section 3 positions this paper to the related work. Section 4 reviews methods and information sources that reveals identifiers used in the network. Revealed identifiers are used to build graphs of identifiers as described in Section 5. Section 5 also defines operations in graphs of identifiers. Section 6 describes the validation of the method. Section 7 concludes the paper.

2 SCENARIOS

This section lists several scenarios that benefit from identifier linking.

Lawful interception (ATIS/TIA, 2006; ETSI, 2009) allows an authorized law enforcement agency (LEA) to capture network traffic of criminal suspects. For each intercept, the LEA provides an identifier that identifies the suspect to a service provider or a network operator. However, the network of the operator can be designed in a way in which it is not straightforward to carry the interception. For example, a lawful interception system has to be able to determine all IP addresses used by devices in the network and link them to the suspect.

Nevertheless, sometimes the warrant authorizing lawful interception provides some restrictions on the identifiers allowed for the interception. For example, a warrant can list an IP address and allow interception of data identified by the particular IP address only. Any other address even if it is certain that the other address belongs to the same computer cannot be intercepted. Other warrants may allow interception of all traffic of the computer. Hence, the linking should be customizable.

Data retention (ETSI, 2015): providers of electronic communications services or public communication networks collect metadata about all communication in the network. Authorized LEA can obtain metadata about communication. A service provider or a network operator that receives a data retention warrant has to find all identifiers used by the suspect during a listed period. Hence, it is crucial that the linking mechanism support time-related queries.

Network forensics entails the separation of suspect traffic from other communication to reduce the amount of information that needs to be analyzed and to avoid privacy infringements of benign users. Nevertheless, as (Casey and Jaquet-Chiffelle, 2017) note, there is always an uncertainty in confidence in the online traces. Malware, identity thefts, and the integrity of the input data are a significant concern. Hence, during network forensic investigation, it is necessary to take into account the confidence in the identification data.

Security incident investigation: Computer emergency response teams (CERTs) and computer security incident response teams (CSIRTs) investigate network security incidents. One of the tasks of the CERT/CSIRT team investigating an incident is to learn the identity of the attacker. For the investigation of the security incident, it can be necessary to link identifiers stored in different sources, sometimes even including end stations.

3 RELATED WORK

Identification in digital environment faces several issues. It is common that a single natural person has multiple online identities (Furnell, 2010; FIDIS project, 2009; Casey and Jaquet-Chiffelle, 2017). Also, identity theft and the use of false identity complicates the identification. The challenge in the current forensic investigation is to prove that the suspect was using the device at the incriminated time, so it is possible to link her responsibility for the actions taken by the device (Jones and Martin, 2010). This paper does not address the linking between natural persons and virtual persons. Nevertheless, the operations allows linking different virtual identities which can be helpful for the identification of the suspect.

(Casey, 2011) emphasizes the need to link identifiers based on specific network parameters and the limits imposed by law. The graphs of identifiers link information revealed from various information sources. Moreover, the operations can limit the scope of allowed linking on a per-query basis. Each query can be limited by time and inaccuracy constraints.

(Da-Yu Kao, 2015) proposes People-Process-Technology-Strategy model to investigate advanced persistent threats, that is stealthy network attacks in which unauthorized persons gain access to a network and remain undetected for a long period. As Da-Yu Kao explains, the technology part of the model benefits from identity linking. Graphs of identifiers help

to achieve the goal to provide tools to execute time-dependent queries.

(Atsa Etoundi Roger and Mboupda Moyo Achille, 2012) describe a network forensic model for proactive, reactive, and active investigation. One of the steps during the forensics is the analysis of gathered evidence. As the evidence is coming from many sources, the analysis can benefit from graphs of identifiers. Nevertheless, expertise supervision on the results of queries in the model is necessary.

Evidence revealed during a forensic investigation has only a limited level of certainty (Casey, 2011; Casey and Jaquet-Chiffelle, 2017). Graphs of identifiers support both accurate and inaccurate relationships. The linking operations consider the inaccuracy and queries can limit linking of inaccurate evidence. Nevertheless, the intended use of the tools is in the early phases of the investigative process (Jackson et al., 2006) when the investigator is looking for leads. The investigator can provide thresholds for acceptable inaccuracy using *linking*[1]. It is up to the investigator to link revealed *partial identities* (Pfitzmann and Hansen, 2010) and *virtual persons* (FIDIS project, 2008a) to the physical world entities.

(Carmagnola et al., 2010; Ye Na et al., 2013; Peled et al., 2013) focus on identity linking of identities in social network websites. This paper provides a method of generalized linking that covers various information sources. Moreover, the linking methods that detect profiles of the same person on different social network websites can be used as another source of relationships between identifiers. The profile identification on a social network is an account identifier of category *L7User* as specified in Section 5.

The main contribution of this paper is the graph model that links identifiers learned from various sources that can be distributed in the network environment. Graphs of identifiers extend previous work introduced by (Polčák et al., 2014b). Graphs of identifiers described in this paper provide multiple advantages compared to the previous work, such as the extensibility, time-related queries, support for identification methods with limited certainty. The enhanced model described in this paper is implemented in the tool *linking*[1].

4 DETECTION OF IDENTIFIERS

Identifiers can be detected from many sources in the network and network hosts. Current research already considers various options to learn identification information. Let us summarize the options.

- Network traffic: network protocols usually rely on identifiers (FIDIS project, 2008b). The presence of a source and destination identifier is often necessary to enable the communication between remote parties. Netflow/IPFIX data provide metadata about the traffic flows in the network.
- Log files, temporary and cache files, and history files reflect events that occurred on a computer system. Events in server logs are typically connected to network-related identifiers (Casey, 2011). See Figure 1 for example of log entries.
- Locally stored information such as hard drives and other storage mediums. Tools such as Autopsy[2] or AUDIT (Karabiyik and Aggarwal, 2014) analyze the content of hard drives including deleted files. Besides the content of the files, metadata, such as the owner, group, or last modification time are also available.
- Hidden identifiers are unique characteristics that are specific to the deployed software or hardware of a unique user, site, or a small set of users. For example, hidden identifiers are browser fingerprints (Laperdrix et al., 2020), communication patterns (Banse et al., 2012; Herrmann et al., 2012; Kirchler et al., 2016), and clock skew (Kohno et al., 2005; Polčák and Franková, 2015) .

(a) *192.168.9.5 - - [04/Nov/2016:15:21:02 +0000] "GET /phpMyAdmin/scripts/setup.php HTTP/1.1" 403 237 "-" "ZmEu"*
(b) *Nov 4 15:06:41 server dhcpd: DHCPACK on 192.168.9.5 to fc:55:47:00:4f:90 (R1) via em0*
(c) *Nov 4 15:22:52 server postfix/smtpd[1234]: Anonymous TLS connection established from dhcp1.example.com[192.168.9.5]: TLSv1.2 with cipher AECDH-AES256-SHA (256/256 bits)*

Figure 1: An example of log files: (a) web server, (b) DHCP server, (c) SMTP server.

Many sources of identification are scattered in the network environment. Even though a single investigation case typically employs only a subset of the discussed identification methods, there is a need to link the scattered information. Graphs of identifiers proposed in Section 5 can utilize any source of identifiers. Each detected relationship between identifiers can be limited in time. Additionally, the identification method should report the uncertainty of the relationship as inaccuracy. The identification method is in the best position to quantify the possible inaccuracy of the provided information as it knows the trustwor-

[2]https://www.sleuthkit.org/autopsy/

thiness of its sources and its detection abilities. Recall that in some scenarios, such as network forensics, all sources have some level of uncertainty depending on the possibility to alter the information (Casey and Jaquet-Chiffelle, 2017).

5 IDENTIFIER LINKING

Section 4 identified various sources of identification information in the network and on network hosts. However, the mere knowledge that some identifiers do exist or did exist at some time is only of limited use for the use cases established in Section 2. This section provides the main contribution of this paper — the model that links the information based on user-specified queries. Subsection 5.1 divides the identifiers into categories according to their similarities, such as durability and the identified subject. Subsection 5.2 defines the graphs of identifiers based on undirected labeled property graphs (Robinson et al., 2015). Finally, Subsection 5.3 defines operations in the graphs of identifiers as implemented by *linking*[1].

5.1 Categories of Identifiers

Even though the identifiers can be learned by various methods, there are some similarities between identifiers. We divide identifiers into categories; each contains identifiers of the same subject — a person, a computer, or a resource. The categories also reflect the occurrence of the identifiers, for example, the identifier appears in each network packet, or the identifier appears during an authentication phase of a protocol only. Last but not least, the categories reflect the duration of the identifiers: long-term or short-term. We consider *virtual identifiers* (FIDIS project, 2008a) only. The goal of the categories is to generalise the similarities in identifiers so that it is possible to define useful operations. The categories are based on categories established by (Polčák et al., 2014b) with two differences: (1) the categories are labeled and (2) we add the category *L7Resource*.

We distinguish the following categories:

L4Flow — Bi-directional TCP and UDP flows.

IPAddr — IP addresses (typically short-term duration and dynamically assigned) identify an interface of a network node.

IfcOrComp — Long-term identifiers of computers or network interfaces, such as MAC addresses, DUIDs, and hidden identifiers, such as clock skew values and browser fingerprints.

AAAUser — Usernames of authentication protocols such as RADIUS, or PPP that identify a set of network devices controlled by a unique virtual person or a household depending on the network.

L7User — Application layer usernames (for example, login names, account identifiers, e-mail addresses) identify a unique virtual person in a specific context or a role. Usually, application layer usernames appear at least once in each session of an application layer protocol. Nevertheless, such session may be composed of several transport layer flows (for example, SIP, FTP).

L7Resource — An application layer resource such as a chat room or a web page URI.

AAAUser and *L7User* identifiers are very close to the natural person whereas *L4Flow* and *IPAddr* identify devices. As the use cases often seek for translation between natural persons and devices or vice versa, such identifiers are common input or output identifiers of the queries in the proposed model.

5.2 Graph Model for Identifier Linking

For identifier linking, we store identifiers in an *undirected labeled property graph* (Robinson et al., 2015). A labeled property graph consists of nodes (vertices) and *relationships* between nodes; each relationship is represented as an oriented edge connecting two nodes. A labeled property graph allows multiple relationships between a pair of nodes. Both nodes and relationships can contain properties as key-value pairs.

In the proposed model, each node (vertex) represents a single identifier. Additionally, each node contains the category of the node as a property. The edges of the proposed model have no orientation.

When an identity source detects a connection between identifiers, the relationship is inserted into the graph of identifiers as an edge. Each relationship is valid during a specific period; the relationship can have an inaccuracy quantified by the detection method. The detection method, validity, and inaccuracy are stored as properties of the relationships in undirected labeled property graphs.

For example, the IP address of a computer of John Doe can be configured through a dynamic protocol such as RADIUS. In that case, the virtual person authenticates a device to the network, and the network assigns the device a dynamic IPv4 address. Such assignment is stored in a RADIUS server log files; it can also be learned from an analysis of RADIUS traffic. In addition, the device might obtain an IPv6 prefix through DHCPv6. Figure 2 shows the graph constructed in this example.

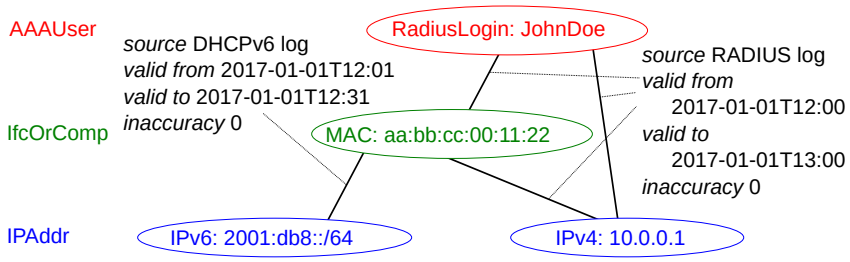

Figure 2: An example of a simple graph of detected identifiers.

5.3 Operations in Graphs of Identifiers

This subsection specifies the operations designed for queries during lawful interception, data retention, network forensics, and security incident investigation. For lawful interception, we define operations reflecting the specific wording of a warrant. Note that it is possible to limit queries in graphs of identifiers by multiple constraints at the same time. A query with multiple constraints yields identifiers that are accepted by all constraints.

The queries are defined as walks in the graph. For mathematical definitions of most of the operations presented in this paper refer to the Ph.D. thesis (Polčák, 2017).

5.3.1 Constraining Relations Between Identifiers

For both network forensics and lawful interception, it is essential to limit identifier linking based on the relationship between the identifiers. For example, (1) Were all linked identifiers used by the same computer? (2) Do all linked identifiers identify the same virtual person?

Based on the restrictions on allowed relationships (edges) between identifiers (nodes), the identity graphs support the following constraints:

Other corresponding identifiers: This constraint aims at cases like a warrant to intercept network traffic linkable to a DHCPv6 DUID. The intercept should cover all IP addresses leased to the DHCPv6 DUID, but it must not cover other IP addresses even if they are assigned to the same interface.

Identifiers of a specific computer: The goal is to detect identifiers belonging to the same computer. For example, a lawful interception warrant may require capturing all traffic of a computer, or, a forensic investigator analyses if two IP addresses belong to the same computer.

Note that the input identifier for this constraint has to be an identifier of a computer or a computer network interface(*IPAddr* or *IfcOrComp*).

Identifiers of computers where a specific virtual person was authenticated or logged in: Sometimes, a lawful interception warrant orders interception of all traffic of all computers authenticated by a specific virtual person or a digital forensic investigator needs such identifiers.

The input identifier of this constraint is an identifier of a virtual person, hence, only *AAAUser* or *L7User* identifiers are allowed.

Identifiers of all virtual persons accessing a specific resource: When network forensics investigator needs to know what users have seen a specific resource, this constrain reveals the information.

All IP addresses accessing a specific resource: For network forensics focusing on a resource, it is usually beneficial to learn all IP addresses that accessed the resource.

All login aliases: A single person can use multiple *L7User* identifiers in parallel. Hence, it is beneficial to have an operation to learn all detected names belonging to a single person — aliases.

All user accounts logged in or authenticated from a computer or a set of computers: The goal is to reveal all virtual person identifiers linkable to an identifier of a computer identified by a category *IPAddr* or *IfcOrComp* identifier, or a set of computers linkable to a category *AAAUser* or *L7User* identifier.

All accessed resources The goal is to determine all resources accessed by a particular virtual person (identified by a category *AAAUser* or *L7User* identifier) or a particular computer (identifier by a category *IPAddr* or *IfcOrComp*) identifier.

5.3.2 Time Restrictions

The time limitations are essential for all considered scenarios. Consider the following constraints:

- The investigator is interested only in identifiers active during a particular period determined by the investigator. Finding such identifiers in a graph of identifiers consists of finding a path between the input identifier and the linked identifier, such as all the relationships (edges) on the path are valid during the whole period.
- The investigator is interested only in identifiers active at any time during a particular period determined by the investigator. In this case, the path consists of relationships (edges) that are valid at some moment of the previous relationship on the path and during the input period.

5.3.3 Path Inaccuracy

In case that some linking information is based on inaccurate sources such as hidden identifiers (see Section 4) or in case of network forensics investigation (Casey and Jaquet-Chiffelle, 2017), the linking is not transitive. Let us define a cumulative inaccuracy for a path in a graph of identifiers as the sum of the inaccuracy of all edges on the path. Then, the query in the graph can be limited with a threshold specifying the maximal inaccuracy of the path from the input identifier to the linked identifier. Another inaccuracy constraints is to ignore all paths with an edge with a cost higher than a threshold.

Note that the model expects that the accuracy is specified by all detection methods with the same metric and that the metric is additive. We leave the definition of a universal metric for future research.

6 VALIDATION

The original mechanism (Polčák et al., 2014b) to link identities is a part of the lawful interception system[3] developed at our university. The extended model proposed in Section 5 released as open source software *linking*[1].

As a part of the lawful interception system, graphs of identifiers were able to link identities that were learned from many sources including DHCP, DHCPv6, SLAAC (and IPv6 neighbor discovery in general), PPPoE, RADIUS, XMPP, OSCAR, IRC, YMSG, and SMTP. Identifiers from these protocols were linked with IP addresses; which in turn were used by custom 1 and 10 Gbps probes to capture traffic of the suspects.

Validation in simulated network Consider an example of a simulated network of an IPv6-enabled internet provider network. The provider authenticates the MAC address of all devices by RADIUS. Each device leases an IPv4 address and generates IPv6 addresses as the device needs (Narten et al., 2007). Figure 3 shows results of a complex forensic investigation in the simulated network. The example considers two monitoring scenarios: local and remote. In the local monitoring scenario, the graph of identifiers is constructed from information in local networks, in this case, RADIUS log files, DHCP log files, and IPv6 neighbor discovery tracking (Polčák et al., 2014a). In a remote scenario, the data in Figure 3 reflects the inability to obtain the data used for local monitoring; instead, the graph of identifiers for remote monitoring is based on inaccurate data sources.

- The number of active IP addresses in the network equals to a sum of the number of IP addresses revealed by learning *other corresponding identifiers* to each RADIUS login at a particular time.
- The average number of IP addresses linked to each active IPv4 address in local monitoring is computed by learning active IPv4 addresses at each evaluation time t_e and the number of linked IP addresses to the input address by learning *identifiers of a specific computer* at the time t_e.
- To present comparable results, the average number of IP addresses linked to each active IPv4 address in remote monitoring is computed by learning active IPv4 addresses at each evaluation time t_e and the number of linked IP addresses evaluating the same query with an inaccuracy thresholds of 3, 5, and 10 for the whole path.

As obvious from the Figure 3, the local and remote monitoring use case provides the forensic investigator with different results. By modifying the inaccuracy threshold, the forensic investigator can focus only on data that most probably belong to the subject of the investigation.

7 CONCLUSION

There are many sources of identifiers in computer networks. One of the challenges is to link information revealed by several sources and methods. The linking is

[3] http://www.fit.vutbr.cz/~ipolcak/prods.php.en?id=397&notitle=1

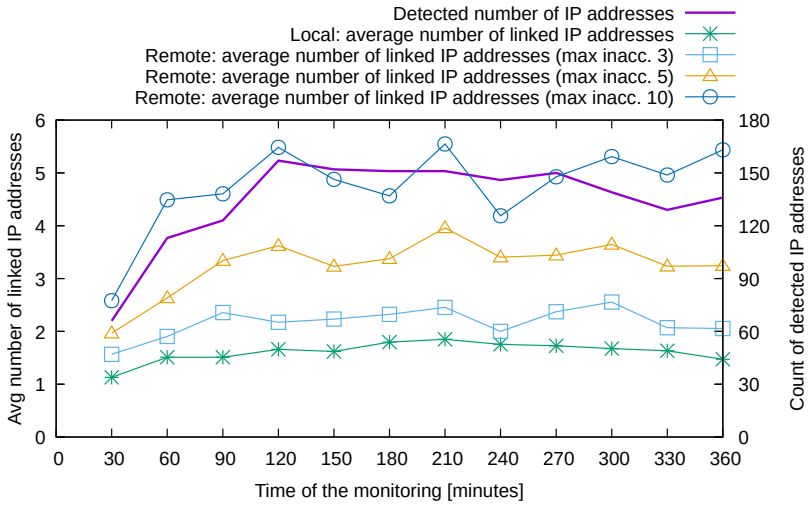

Figure 3: The number of detected IP addresses (right y axis) and linked average number of IP addresses (left y axis) in the simulated internet service provider network.

applicable in (1) lawful interception as a lawful interception system has to identify traffic of an interception target according to the warrant allowing the intercept, (2) data retention queries, (3) early stages of network forensic investigative process, and (4) security incident investigations.

This paper describes graphs of identifiers that extend the work of (Polčák et al., 2014b). Graphs of identifiers can be built based on information from various sources, including traffic traces, log files, and inaccurate identification methods. One of the essential extensions to the previous work is the time that is an inherent part of graphs of identifiers. Consequently, the investigation queries support time-related queries. Another significant extension is the support of inaccurate identification methods which quantify the inaccuracy for each reported relation between two identifiers.

The tool *linking*[1] is freely available and implements queries in graphs of identifiers as described in this paper. We also provide an extensible log files converter that currently allows processing log files of ISC DHCP server and NCSA common/combined log format.

ACKNOWLEDGEMENTS

This work was supported by The Ministry of Education, Youth and Sports of the Czech Republic from the National Programme of Sustainability (NPU II); project IT4Innovations excellence in science - LQ1602. We thank Frank Breitinger for his help during the preparation of this paper.

REFERENCES

ATIS/TIA (2006). *Lawfully Authorized Electronic Surveillance. J-STD-025-B*. Alliance for Telecommunications Industry Solutions/Telecommunications Industry Association Joint Standard.

Atsa Etoundi Roger and Mboupda Moyo Achille (2012). Multi-perspective cybercrime investigation process modeling. *International Journal of Applied Information Systems*, 2(2):14–20.

Banse, C., Herrmann, D., and Federrath, H. (2012). Tracking users on the internet with behavioral patterns: Evaluation of its practical feasibility. In *Information Security and Privacy Research*, volume 376 of *IFIP Advances in Information and Communication Technology*, pages 235–248. Springer Berlin Heidelberg, DE.

Carmagnola, F., Osborne, F., and Torre, I. (2010). User data distributed on the social web: How to identify users on different social systems and collecting data about them. In *Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, HetRec '10, pages 9–15, New York, NY, USA. ACM.

Casey, E. (2011). *Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet*. Academic Press, Elsevier Inc., USA. Third Edition.

Casey, E. and Jaquet-Chiffelle, D.-O. (2017). Do identities matter? *Policing: a Journal of Policy and Practice*, (Special Issue):1–14.

Da-Yu Kao (2015). Performing an APT investigation: Using people-process-technology-strategy model in digital triage forensics. In *2015 IEEE 39th Annual Computer Software and Applications Conference*, volume 3, pages 47–52.

ETSI (2009). *ETSI TR 101 331: Lawful Interception (LI); Requirements of Law Enforcement Agencies*. Euro-

pean Telecommunications Standards Institute. Version 1.3.1.

ETSI (2015). *ETSI TS 102 657: Lawful Interception (LI); Retained data handling; Handover interface for the request and delivery of retained data*. European Telecommunications Standards Institute. Version 1.17.1.

FIDIS project (2008a). D2.13: Virtual Persons and Identities. David-Oliver Jaquet-Chiffelle (ed.), Version 1.0, Available online at http://www.fidis.net/resources/fidis-deliverables/identity-of-identity/#c2162.

FIDIS project (2008b). D3.8: Study on protocols with respect to identity and identification — an insight on network protocols and privacy-aware communication. Marit Hansen and Ammar Alkassar (ed.), Version 0.8. Available online at http://www.fidis.net/resources/fidis-deliverables/hightechid/#c2216.

FIDIS project (2009). D17.4: Trust and Identification in the Light of Virtual Persons. David-Oliver Jaquet-Chiffelle and Hans Buitelaar (ed.), Version 1.2. Available online at http://www.fidis.net/resources/fidis-deliverables/identity-of-identity/#c2596.

Furnell, S. (2010). Online identity: Giving it all away? *Information Security Technical Report*, 15(2):42–46. Identity Theft and Reconstruction.

Herrmann, D., Gerber, C., Banse, C., and Federrath, H. (2012). Analyzing characteristic host access patterns for re-identification of web user sessions. In *Information Security Technology for Applications*, volume 7127 of *Lecture Notes in Computer Science*, pages 136–154. Springer Berlin Heidelberg, DE.

Jackson, G., Jones, S., Booth, G., Champod, C., and Evett, I. (2006). The nature of forensic science opinion — a possible framework to guide thinking and practicce in investigation and in court proceedings. *Science & Justice*, 46(1):33–44.

Jones, A. and Martin, T. (2010). Digital forensics and the issues of identity. *Information Security Technical Report*, 15(2):67–71. Identity Theft and Reconstruction.

Karabiyik, U. and Aggarwal, S. (2014). Audit: Automated disk investigation toolkit. *The Journal of Digital Forensics, Security and Law*, 2014(2):129–143.

Kirchler, M., Herrmann, D., Lindemann, J., and Kloft, M. (2016). Tracked without a trace: Linking sessions of users by unsupervised learning of patterns in their DNS traffic. In *Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security*, AISec '16, pages 23–34, New York, NY, USA. ACM.

Kohno, T., Broido, A., and Claffy, K. C. (2005). Remote physical device fingerprinting. *IEEE Transactions on Dependable and Secure Computing*, 2(2):93–108.

Laperdrix, P., Bielova, N., Baudry, B., and Avoine, G. (2020). Browser fingerprinting: A survey. *ACM Trans. Web*, 14(2):8:1–8:33.

Narten, T., Draves, R., and Krishnan, S. (2007). *Privacy Extensions for Stateless Address Autoconfiguration in IPv6*. IETF. RFC 4941 (Draft Standard).

Peled, O., Fire, M., Rokach, L., and Elovici, Y. (2013). Entity matching in online social networks. In *2013 International Conference on Social Computing*, pages 339–344.

Pfitzmann, A. and Hansen, M. (2010). A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. Technical report. Version 0.34, Available online at https://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf.

Polčák, L. (2017). *Lawful Interception: Identity Detection*. PhD thesis, Brno University of Technology, Faculty of Information Technology.

Polčák, L. and Franková, B. (2015). Clock-skew-based computer identification: Traps and pitfalls. *Journal of Universal Computer Science*, 21(9):1210–1233.

Polčák, L., Holkovič, M., and Matoušek, P. (2014a). Host Identity Detection in IPv6 Networks. In *Communications in Computer and Information Science*. Springer Berlin Heidelberg, DE.

Polčák, L., Hranický, R., and Martínek, T. (2014b). On identities in modern networks. *The Journal of Digital Forensics, Security and Law*, 2014(2):9–22.

Robinson, I., Webber, J., and Eifrem, E. (2015). *Graph Databases*. O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA, USA. Second Edition.

Ye Na, Zhao Yinliang, Dong Lili, Bian Genqing, Liu, E., and Clapworthy, G. J. (2013). User identification based on multiple attribute decision making in social networks. *China Communications*, 10(12):37–49.