Challenges in Identification in Future Computer Networks

Libor Polčák
Faculty of Information Technology, Brno University of Technology, Božetěchova 2, 612 66 Brno, Czech Republic

1 OUTLINE OF OBJECTIVES

The knowledge of user identity is a necessity for many network-related tasks, e.g. network management (Grégr et al., 2011), security incident backtracking (Scarfone et al., 2008), authentication and authorisation (Huang et al., 2012; Sanguanpong and KohtArsa, 2013), lawful interception (ATIS/TIA, 2006; ETSI, 2001), and quality of service (Laiping Zhao et al., 2012). Traditionally, an IPv4 address was a suitable identifier for the identification. With the advent of new technologies, such as IPv6 or *carrier grade network address translation* (CGN/NAT444), the methods for user identification need to be revised (Polčák et al., 2013a).

In IPv4, a device usually leases an IPv4 address from a DHCP server in the custody of the network operator. Then, the device applies this unique IPv4 address for all its communication until the lease expires. In contrast, IPv6 introduced several new mechanisms for address assignments. For example, *Stateless Address Autoconfiguration* (SLAAC) (Thomson et al., 2007) allows an end device to generate as many IPv6 addresses as it needs, e.g. for privacy concerns (Groat et al., 2010; Narten et al., 2007), as long as the addresses are not already used by another device in the network. Note that the addresses are not handled centrally but generated by end devices.

Some companies prefer to prioritize specific traffic, e.g. important video calls of a manager with high bandwidth demand. Lately, companies allow their employees to bring their own equipment to work (bring your own device policy). These devices are not registered in the network, consequently, their identity is unknown. Hence, a device of a manager is indistinguishable from other devices; and the traffic of such an equipment cannot be treated according to special rules easily.

This Ph.D. research focuses on user identification in future computer networks. The aim is to study the information available in different parts of the network (e.g. local area networks — LANs, backbone networks, content provider networks etc.). The ultimate goal is to propose mechanisms that identify the user even though the user changes his or her IP address. Although the primary aim is at identification of traffic of a specific user, identification of data of a specific computer is also considered since it might ease the primary goal of user identification. In addition, the research takes into account latest network technologies: CGN/NAT444, IPv6 and software defined networking (SDN) are considered during the research.

The research is divided into the following areas:

- A study and improvements of existing methods for identification or proposal of new methods. In addition, the research should evaluate advantages and disadvantages of the methods. The aim is at methods that does not require cooperation from the end user and are transparent for him or her. These methods are applicable to all of our use cases.
- Understanding of relations between different identities detected on different layers of the TCP/IP architecture. The Ph.D. research should distinguish between identities of computers and persons.
- Proposal of mechanisms to link identities of the same person or a computer.
- Proof-of-concept: results of the Ph.D. research are expected to be used as a part of the Lawful Interception System developed at the Brno University of Technology as a part of the research focused on criminality mitigation on the new generation Internet (FIT BUT, 2014). Additionally, we plan to develop an SDN application that controls an SDN network and prioritize traffic according to the identity of the communicating parties.

This section outlined the objectives of the Ph.D. research. Section 2 introduces the terminology and

Polčák L..
Challenges in Identification in Future Computer Networks.

explains the research problem in detail. Section 3 overviews the work related to this Ph.D. research. The methodology to achieve the goals is proposed in Section 4. Section 5 elaborates on the expected outcome of the Ph.D. research and portrays the validation of the research. Current stage of the research is discussed in Section 6 and the contribution is summarized in Section 7.

2 RESEARCH PROBLEM

In conformance with (Pfitzmann and Hansen, 2010), we consider an *identity* to be a set of *attribute values* that uniquely identify a subject, i.e. a person or a computer.

A single identity may comprise (Pfitzmann and Hansen, 2010) of different *partial identities* of the same person or a computer; each of the partial identity represent the subject in a specific context or a role. All partial identities co-exist and can potentially be *linked* (correlated). If two identities are *linkable*, their attributes can be merged as all belong to the same subject. Consider a user owning two e-mail addresses — x any y. Both x and y are partial identities of the subject.

In the network environment, often, one of the attributes is unique for a specific identity. Such attribute is called the *identifier* — usually it exists in a form of a name or a bit string (Pfitzmann and Hansen, 2010). Each identifier corresponds to a specific (partial) identity. For instance, the identity of an e-mail user owning a mail box of an address *x* can be represented directly by the identifier *x*. This Ph.D. research should focus on the identifiers and how common they are.

Note that for privacy reason, we are not interested in gathering extensive amount of attributes and their values during this Ph.D. research. For quality of service, we are primarily interested in the network identifiers so that it is possible to distinguish network traffic of different users. The lawful interception use case has to follow strict legal rules that prevents pervasive monitoring of all users. Hence, the goal is to learn the identifiers, link the partial identities represented by each of them and identify the traffic of the discovered subject.

The problem of user or computer identity detection is useful in several domains, however, its complexity slightly differs. The basic difference is the availability of specific attributes in a specific case: some of the sources of attributes are not always available, or, some sources cannot be utilized for a reason, e.g. they are not reliable enough for that case. For this research, we restricted the sources of attributes to those that are transparent for the user to be identified. Although this restriction limits the methods that are available, it is a necessary restriction because we are interested in methods that are compatible with *lawful interception* (LI) (ETSI, 2001). However, the considered method to link partial identities is general and it does not restrict any method that can represent an identity by an identifier. Hence the method is compatible with various techniques that discover partial identities, including those that are not transparent to the end user.

The research has to cover several scenarios for the identification. The goal is to identify the traffic of end users in the network. We aim to quantify the usability of the methods that have been already proposed, on their improvements, and on development of new methods. Based on the use cases of an application aware network and the deployment of a *lawful interception system* (LIS), we consider identification in LANs and remote networks. In addition, we consider modern technologies, such as IPv6 and software defined networking.

To achieve this goal of traffic identification, the research needs to study the attributes, identities, relations between identities, and linkability of partial identities. The ultimate goal is to present methods suitable for various kinds of modern and future networks, including:

- Network address translation (NAT), carrier grade NAT (CGN) or multi-layer NAT (NAT444): As depicted in Figure 1, IPv4 addresses are shared between several households while one computer can communicate through different translators that use different IPv4 addresses.

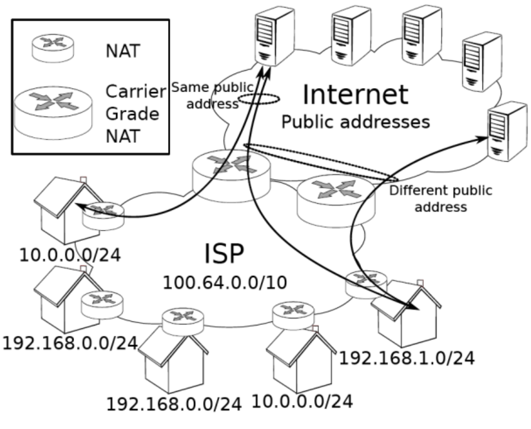

Figure 1: Multi-layered translation of IPv4 addresses.

- IPv6 networks: Short-lived temporary IPv6 addresses (Narten et al., 2007) can be generated by any computer in IPv6 network at will. Moreover, a computer can use as many IPv6 addresses on each interface as it can handle. Furthermore, recent versions of Windows, Mac OS X, iOS, and several Linux distributions have temporary addresses enabled by default. Usually, a new temporary address is generated at least once per day. However, when a user authenticates with a different access point in a Wi-Fi network or reboots his or her computer, it regenerates its temporary addresses. Figure 2 portrays default behaviour of a Windows computer.

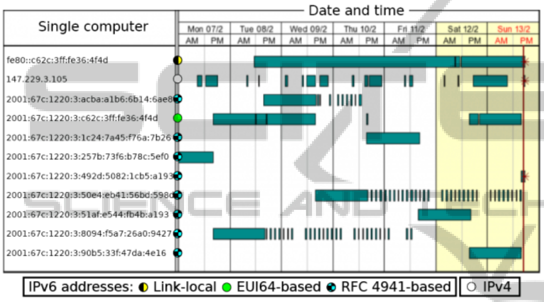

Figure 2: Multiple IPv6 addresses were used during a week by a single Windows-based computer simultaneously. The computer was not shut down during that week. While it used only one IPv4 address, it generated new IPv6 address every day.

  Hence, IPv6 hosts use several identifiers for the same interface. Consequently, there needs to be a mechanism to link all addresses of one computer together.

- Dual stack networks: Usually, even in IPv6enabled networks, IPv4 is still present (as also illustrated in Figure 2). Recent operating systems and web browsers employ Happy eyeballs (HE) (Wing and Yourtchenko, 2012) to dynamically select the protocol with better performance (sometimes with a slight preference of IPv6). As a result, one session (e.g. web session) can be split between both protocols.

  In addition, even without HE, dual-stacked machines communicate with IPv4-only Internet using IPv4 while IPv6-enabled servers are accessed via IPv6. As web pages often contain external content and DNS is accessed separately, one session may be split between IPv4 and IPv6 even without HE.

3 STATE OF THE ART

This Section outlines the related work to this Ph.D. research. First, the focus is on the terminology and other approaches to identification. Then, the focus shifts on detection methods that are not detectable by an end user of a network. Several methods suitable for passive user identity detection already emerged: the characteristics of a subject (attributes) are enclosed in the data that the subject exchanges in the network or in the metadata of the communication.

3.1 Identity Management Systems

FIDIS was an European project that focused on identification. They considered (Meints and Gasson, 2009) three types of *identity management systems* (IMSes):

1. IMSes for account management, *authentic-cation, authorisation, and accounting* (AAA).
2. IMSes for profiling of user data by an organisation, for marketing or providing personalised services.
3. IMSes for pseudonym management. These systems are controlled by the end users.

Indeed, this Ph.D. research concerns IMSes. Our IMS can be characterised as a mix of type 1 and 2. The aim of our IMS is to link different identities from computer networks. The ultimate goal is to provide personalised services for all traffic of a specific user or a group of users. As the sources of identities are not limited, it is possible to use any Type 1 IMS as a source of identities for our IMS. Therefore, our IMS can be used as Type 1 IMS and provide authorisation or accounting for related identities. Such scheme can be deployed on dual stacked networks (Sanguanpong and Koht-Arsa, 2013) where many network-layer addresses are used at the same time.

There is an extensive research in identity area. For example, (Jøsang et al., 2005) list relations between identifiers, identities and entities. They explain that one entity has several (partial) identities, each of them can be identified by several identifiers. As they focused on a type 3 IMS, they did not pursued the idea of linking identities as this research does. Compared to their research, this research aims to reveal different identities of the same person.

Compared to definitions of (Clauß and Köhntopp, 2001), we are not focused on *gathering huge amount of data*. Our goal is to identify a person and its network traffic. Their notion of transaction, situation, and person pseudonyms might be proven useful during the formal specification of mechanisms to link identities.

3.2 Identity Detection

The metadata-based approach to reveal identities can be used for both local and remote monitoring. One of the possibilities are behaviour models (Banse et al., 2012; Herrmann et al., 2012; Kumpošt, 2008). The models are typically constructed of metadata of past communication, e.g. accessed IP addresses, amount of exchanged data, time in a day of communications with specific host, etc. The disadvantage of the models is the need to track the user for a long time, e.g. one day (Banse et al., 2012), and often the models have difficulties with users regularly changing their IP addresses (Herrmann et al., 2012). In contrast, common operating systems use temporary IPv6 addresses by default; a typical preferred life time of a single IPv6 address is one day.

HTTP headers and information learnt from JavaScript were originally studied for active identification (Eckersley, 2010). However, a stable set of HTTP headers (such as a user agent string and similar headers) is present in all HTTP requests of a single browser instance. Consequently, the set can be treated as additional attributes used for passive identification. The usability of the method varies with the amount of information that a browser sends to the Internet and it needs to be further investigated. Since browser version is one of the attributes monitored by the method, the fingerprint of a browser changes with every update.

Kohno et al. (Kohno et al., 2005) proposed identification of computers by measuring their clock skew computed from ICMP and TCP timestamps. Later, new sources of time information were proposed (Huang et al., 2012; Murdoch, 2006; Zander and Murdoch, 2008; Lanze et al., 2012). The advantage of clock-skew-based identification lays in its speed (Huang et al., 2012; Sharma et al., 2012) and the method also works for mobile devices (Sharma et al., 2012). The possible sources have some limitations:

- Windows clients does not send TCP timestamps (Kohno et al., 2005).
- ICMP timestamps are not supported by Apple (Polčák and Franková, 2014) and this source is not available in IPv6.
- Timestamps from application protocols have either low resolution (Zander and Murdoch, 2008) or are not present if not requested by a server (Huang et al., 2012).

However, the advantage of the clock-skew-based identification is that the observed clock skew does not change with an update of the software, after a change of a user behaviour, after a relocation to another network, or after a switch between wired and wireless network.

Several attempts have been made to study IPv6 addresses and their assignments. Static interface IDs (lower 64-bits of an IPv6 address) were proposed (Groat et al., 2010; Dunlop et al., 2011) as a mean to monitor the location of a roaming node. The downsides of the method are twofold: (1) the interface ID needs to be known in advance and (2) *ping* or *tracer-oute* probe packets are in the worst case sent to all networks in the Internet. To limit the amount of probe packets, the method needs to be focused on a specific set of usual locations of the tracked user. This Ph.D. research focuses on passive monitoring, and therefore, this method is not applicable.

Grégr et al. (Grégr et al., 2011) poll *neighbor cache* of the routers in our University network to gather information about address assignments. However, the gathered MAC addresses are not available outside of the local network, and consequently, this method is not applicable for the remote location tracking scenario. Another downside is that the polling of the routers causes additional workload on the routers.

Groat et al. (Groat et al., 2011) studied DHCPv6 for monitoring the identity of users in LANs. However, they focused only on one specific address assignment method — DHCPv6. In contrast, SLAAC is a default address assignment method in IPv6 and DHCPv6 is deployed only rarely. Moreover, even if it is deployed, end hosts can still employ SLAAC to generate additional IPv6 addresses.

Asati and Wing (Asati and Wing, 2012) tried to propagate the information about address assignments outside of local networks but their effort did not make it through the IETF standardisation process.

3.3 Proprietary Solutions

Napatech[1] summarized their thoughts on problems in current networking in a series of white papers. Similarly to our research, (Napatech, 2014a) calls for a distributed solution with deep packet

[1] http://www.napatech.com

inspection capabilities and application detection. These methods might be employed as a source of identities in our approach. The processing and correlation of the information gathered in their white paper is not clear. Napatech calls for better knowledge of customer needs and behaviour to avoid death spiral of decreasing revenue. The mechanism to link (correlate) identities developed during this research should be capable to help in this use case as it can derive the identities of the same user running different applications. In addition, (Napatech, 2014b) advocates for analysis of all network layers. This is very similar to our approach. The downside of the white papers is the omission of the detailed description due to their solutions being proprietary.

Cisco Medianet Metadata (Cisco Systems, 2014) allows to treat flows of specific applications according to category, their urgency for business, etc. Similarly, the framework for signaling of flow characteristics (Eckert et al., 2013) tags flows with metadata. Both solutions interact with an application running on an end host and passes metadata about flows of the application to the network environment. Hence, both can be used as a source of partial identities.

4 METHODOLOGY

Since there is not a single method that deals with all identification-related challenges, several approaches need to be considered. Typically, sources of information about identities are scattered in the network. Moreover, each source has only limited knowledge about the attributes and consequently, it operates only with partial identities.

This Section focuses on identities, their detection and processing. Firstly, it elaborates on the linkability of identities. Later, it outlines the possible sources of information that are investigated during this Ph.D. research.

4.1 Linkability of Identities

We consider a distributed solution to discover the identities as depicted in the Figure 3. All sources of identities present a different view on the network. For example, *RADIUS* can reveal logins of the users in the network but IPv6 addresses has to be discovered from other sources, such as *IPv6-SLAAC* or *DHCPv6*. The holistic view on the identities is achieved in the central device by correlation of all information leant from the sources.

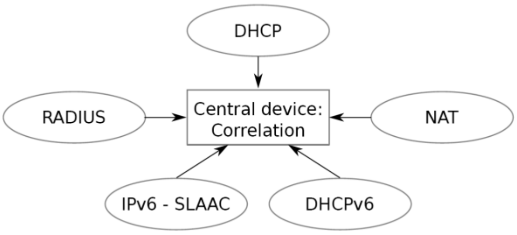

Figure 3: Identities can be learnt from different sources in the network and correlated in the central device.

The sources of discovered identities reveal network identifiers that can be used in place of the represented identity. For example, the e-mail address x can represent the owner of the mail box. Typically, basic relations between partial identities can be uncovered from the network sources, such as *the owner of the e-mail address x connected to its mail box from a computer with the IP address y*.

Usually, one or more identifiers are common between several sources of identities; they can be used to reveal additional relations between partial identities. Consider a user authenticating through RADIUS. He or she authenticates their MAC addresses when they access a network. Later, their computer generates IPv6 addresses with SLAAC. When both mechanisms are monitored, the relation between IPv6 addresses and the RADIUS login can be revealed through the mutually discovered MAC address.

We consider graph representation of the discovered (partial) identities. The constructed graph has to be able to express different types of network identifiers, from all layers of the TCP/IP architecture. In addition, NAT and its implication on the graph structure has to be considered. Hence, the aim of the research should be at better understanding of network user identities, their relations and the identifiers that sufficiently represent the identities.

The graph depicted in Figure 4 represents a snapshot of identities discovered in a monitored network of a company. A user is authenticated to the network and his or her computer is using an IPv4 and an IPv6 address. The traffic of these IP addresses can be handled by special rules, e.g. it should be treated with higher priority since the user happens to be the manager of the company.

We see benefits in graph representation of the snapshots of identities detected in a network at a specific time. Indeed, a graph can model relations between identities. Moreover, algorithms that are wellknown from graph theory can be used to infer additional information from the graph, such as

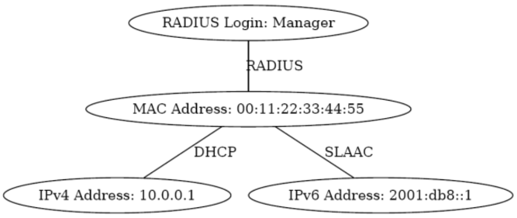

Figure 4: An example of the graph representing relations between different identities discovered in a network. The identities are represented by network identifiers depicted as nodes. The source of the relation between different partial identities is displayed as a label of each edge.

related partial identities of the same subject. Another goal is to represent graphs so that identities related to computers and identities related to specific persons can be distinguished.

Consequent plans include incorporating time information into the graphs. Each change in the network can be considered to be a shift between two states of the network. Each state of the network can be represented with the above-mentioned graph. This is useful for the network monitoring use case.

The rest of this Section focuses on the sources of identities in the network.

4.2 Local Monitoring

In local monitoring, local knowledge of the network can be utilised. Often, an application manages the users in the network, their computers or available services: for instance a DHCP server or a RADIUS server. Additionally, consider application layer services, e.g. instant messaging server, VoIP private branch exchange or SMTP server, as another example. All these services are instances of the *Type 1 identity management system* (IMSt1) (Meints and Gasson, 2009).

Such IMSt1s are the most precise sources about the identities connected to the network. Identities can be learnt either in direct cooperation with an IMSt1 or from the outputs of these IMSt1s, e.g. logs.

a) Figure 5 depicts an extension or a plug-in of an IMSt1. The plug-in is specialized to advertise the managed identities. When a new identity is learnt by the IMSt1, the plug-in immediately passes the information and the identity can be later correlated with other identities from other sources. Since this approach has to be supported by the IMSt1, it is not applicable in all cases: for instance, a plug-in system is not available, or, the IMSt1 cannot be accessed by the monitoring entity. Cisco Medianet Metadata (Cisco Systems, 2014) and the framework for signaling of flow characteristics (Eckert et al., 2013) are examples of this scenario whenever the IMSt1 signalise attributes and identifiers of flows through these channels.

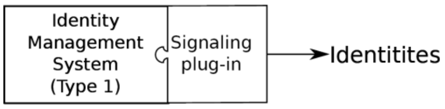

Figure 5: Direct access to an IMSt1 allows immediate access to all managed identities by the IMSt1.

b) IMSt1s often log changes of the state to external logs, e.g. stored on a local hard drive. These logs can be regularly analysed by a script, which discovers attributes and consequently identities in the logs (as displayed in Figure 6). In this case, the changes are learnt with a slight delay caused by the polling. The delay can be avoided in case the file system allows the script to be notified after the log was changed, e.g. through *inotify* (Love, 2005). Alternatively, SNMP traps or similar mechanisms can be used when logs are not available.

However, sometimes a special code cannot be integrated to an IMSt1 of interest, the logs are not available, or the delay caused be polling of the logs is not acceptable. For these circumstances, we focus on traffic parsing during the Ph.D. research.

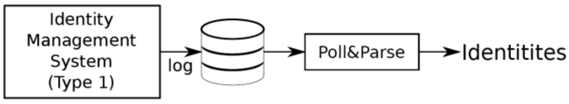

Figure 6: Logs of IMSt1s typically contain information about the processed identities, which can be collected by a script and passed for a further processing.

In the case of centralised services, such as DHCP or RADIUS, we propose to gather the information as close to the server as possible; preferably on the access link of the server. In this way, all traffic destined for the server can be analysed with one probe.

However, addresses generated during SLAAC are not handled centrally but the knowledge about the assignments is distributed in the network (Polčák et al., 2013a). For this case, we already proposed (Polčák et al., 2013a) tracking of IPv6 control traffic (ICMPv6). The approach is passive and it can learn the assignments of all IPv6 addresses in a LAN.

In SDN, a controller or a cluster of controllers have a detailed knowledge (McKeown et al., 2008) about the network including its topology and end hosts. Usually, the controller has a *northbound interface* that can be used by a third party application to learn about the state of the network or inject rules. Therefore, an SDN controller can be another source of information about the identities in a network.

Today, typically all network interfaces are identifiable by a MAC address. Since the ICMPv6 approach (Polčák et al., 2013a), DHCP logs, and SDN controllers provide MAC and IP address pairings, different identities of one computer can be linked through MAC addresses. However, note that not all mechanisms can reveal a MAC address, for instance DHCPv6 (Polčák et al., 2013a) does not assign IPv6 addresses according to MAC addresses but instead, it uses DUIDs — special identifiers of DHCPv6.

In summary, various mechanisms can be used to detect identities in local networks. A direct cooperation with IMSt1s provides most precise information, however, it is not always available. SDN controllers are another source of reliable information. However, some identities can be only learnt from traffic parsing. This research consider all these means to obtain identities.

4.3 Remote Monitoring

When identities of remote users need to be detected, e.g. for an Internet access provider monitoring its network or guaranteeing quality of service, Layer 2 and lower identifiers are not available during remote identity detection, unless a MAC address leaks, e.g. through lower part of an IPv6 address (Dunlop et al., 2011). As such leaks are not common, a different method to link identities needs to be employed, instead.

Each computer has internal clock to measure time. As the manufacturing process is not precise on atomic level, each clock has its own deficiencies. Consequently, each computer measures time with its own in-built inaccuracy, *clock skew*. Since the clock skew does not depend on the location of a computer in the network, or the type of its connection (copper or WiFi), a computer can be identified (Kohno et al., 2005) as it moves from network to network. To evaluate this possibility, a part of the Ph.D. research focuses on clock skews and their suitability for remote identification.

HTTP headers also reveal substantial information (Eckersley, 2010) about user identity. Besides the headers, Eckersley also considered JavaScipt and cookies. However, for passive detection, only HTTP headers from requests are available.

Additionally, clock skew and the browser fingerprinting based on HTTP headers can be combined. Therefore, even computers with very similar clock skew can be distinguished by potentially different HTTP headers. Vice versa, browsers with similar configuration can be running on computers with different clock skew; thus the browsers can be differentiated through the clock skew. In addition, browser updates can be linked through the detection of a known clock skew.

The behaviour models (Banse et al., 2012; Herrmann et al., 2012; Kumpošt, 2008) are not considered during this Ph.D. research. As mentioned in Section 3, the reported recognition time of a computer is one day or longer. As users roam between networks much quicker, the behaviour models are not suitable for this Ph.D. research.

5 EXPECTED OUTCOME

This section elaborates on the content of the final Ph.D. thesis. Firstly, the thesis has to list the challenges for identification in modern networks. This list should be based on the challenges solved during the research and open challenges for future research.

Additionally, the Ph.D. thesis should provide formal algorithms and methodology for dealing with partial identities and their linkability. The algorithms need to support various sources of information and they have to consider possible network operational conditions, including NAT and multiple addresses of the same computer.

Previously proposed solutions should be evaluated as the sources of information. Moreover, the Ph.D. thesis should introduce new mechanisms for identity detection. All considered mechanisms should be examined and their advantages and disadvantages has to be investigated.

In addition, the thesis should reflect possible use cases related to identity detection. Lawful interception, network management, and provisioning of high quality of service are among the possible applications of this work. As such, the proposed solution needs to be tested for one of the above use cases.

As one of the aims of the research carried at the Brno University of Technology is focused on criminality mitigation on the new generation Internet (FIT BUT, 2014), incorporation of the proposed mechanisms to the Lawful Interception System is an obvious choice.

We also plan development of an SDN-based management system for traffic shaping and controlling network accessibility based on identities. For example, consider an IT company with a development department, HR, and accounting. Each of these departments has servers and services that should not be accessed by employees of other departments. The IMS designed during this Ph.D. research can differentiate the traffic of employees of different departments and provide separation even if employees of all departments are in one room during a meeting.

6 STAGE OF THE RESEARCH

This section lists the achievements that were accomplished during the research. The research already achieved several milestones both in identity management and in studying the sources of partial identities.

In the area of local monitoring, we proposed a mechanism to detect all IPv6 addresses used in the network by monitoring messages exchanged during *neighbor discovery*. The method takes into account several differences that we discovered (Polčák and Holkovič, 2013) in the behaviour of operating systems. This method was successfully tested at our University network and it can reveal all IPv6 addresses used by one computer. Later, we expanded (Polčák et al., 2013b) the detection to SDN environments.

In remote monitoring, we were interested in the clock-skew-based identification (Kohno et al., 2005). The method looked appealing because the clock skew does not depend on computer interface or its location. We were interested in using the method to link different IPv6 addresses of the same computer. This worked in laboratory environment, however, our measurements revealed several obstacles (Polčák et al., 2013c; Polčák and Franková, 2014) that makes the method hard to use in real networks. We consider combining the clock-skew-based identification with other methods, such as those based on unique content of HTTP headers (Eckersley, 2010) or amount of traffic sent by specific users (Megyesi and Molnár, 2013).

We already proposed a formal method to link identities but it was not published, yet. Currently, we wait for the results of the peer review for the submitted paper concerning the approach.

All methods have already been incorporated to the lawful interception system developed at Brno University of Technology (FIT BUT, 2014). We plan to show the application of the method in SDN with the aim of provisioning desired quality of service and controlling access to specific parts of network based on the identity of a user.

7 CONCLUSION

Modern computer networks bring new challenges, such as network address translation and short-lived IPv6 addresses. As a result the number of identifiers related to a single user increases. Consequently, old methods for identification of the traffic of a specific user are becoming weak. This research aims to tackle the challenges by linking partial identities of a subject together. This way, all traffic of specific groups of users can be identified and it can be treated in a personalised manner, e.g. important calls of a manager can be prioritised.

We consider a distributed system that can link identities discovered from various sources. As an example, we investigated both local and remote identification. This research already resulted in several accepted papers (Polčák et al., 2013a; Polčák et al., 2013b; Polčák et al., 2013c; Polčák and Franková, 2014). Current efforts are in the area of the proposal of the method to link different identities, in improvements of the remote identification techniques, and in designing an SDN-based control system expanding the proposed IMS to network control, such as quality of service.

ACKNOWLEDGEMENTS

This research belongs to the project VG20102015022 (Modern Tools for Detection and Mitigation of Cyber Criminality on the New Generation Internet) supported by the Ministry of the Interior of the Czech Republic. It was also supported by the project FIT-S-14-2299 (Research and application of advanced methods in ICT) of Brno University of Technology.

I would like to thank my supervisors Miroslav Švéda and Petr Matoušek for their help during the research and for their input. In addition, I would like to highlight the contribution of Tomáš Martínek who helped me with this research during our discussions held on our regular meetings. His ideas and comments improved this work. I would also like to thank my students, especially Stanislav Bárta, Barbora Franková, Martin Holkovič, Radek Hranický, Jakub Jirásek, and Petr Kramoliš: both for their input and for their help during the development of the tools required in this research.

Last but not least, I want to thank Cisco Systems Switzerland, in particular Davide Cuda and Amine Choukir, for giving me the possibility of seeing the research problem from a different angle.

REFERENCES

Asati, R. and Wing, D. (2012). *Internet Draft version 00 (Work in progress): Tracking of Static/ Autoconfigured IPv6 addresses*. Internet Engineering Task Force.

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

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

Cisco Systems (2014). Cisco medianet architecture. http:// www.cisco.com/web/solutions/trends/medianet.

Clauß, S. and Köhntopp, M. (2001). Identity management and its support of multilateral security. *Computer Networks*, 37(2):205–219. Electronic Business Systems.

Dunlop, M., Groat, S., Marchany, R., and Tront, J. (2011). The Good, the Bad, the IPv6. In *Communication Networks and Services Research Conference*, pages 77–84, Ottawa, Canada.

Eckersley, P. (2010). How unique is your web browser? In Atallah, M. and Hopper, N., editors, *Privacy Enhancing Technologies*, volume 6205 of *Lecture Notes in Computer Science*, pages 1–18. Springer Berlin Heidelberg, DE.

Eckert, T., Penno, R., Choukir, A., and Eckel, C. (2013). *A Framework for Signaling Flow Characteristics between Applications and the Network*. IETF. InternetDraft (work in progress), intended status: Informational.

ETSI (2001). ETSI TR 101 944: *Telecommunications security; Lawful Interception (LI); Issues on IP Interception*. European Telecommunications Standards Institute. Version 1.1.2.

FIT BUT (2010–2014). Modern tools for detection and mitigation of cyber criminality on the new generation internet. Brno University of Technology, Faculty of Information Technology, CZ. http://www.fit.vutbr.cz/~ipolcak/ grants.php?id=517.

Groat, S., Dunlop, M., Marchany, R., and Tront, J. (2010). The privacy implications of stateless IPv6 addressing. In Cyber Security and Information Intelligence Research, pages 52:1–52:4, New York, NY, USA. ACM.

Groat, S., Dunlop, M., Marchany, R., and Tront, J. (2011). What DHCPv6 says about you. In 2011 World Congress on Internet Security, pages 146–151, London, UK.

Grégr, M., Matoušek, P., Podermański, T., and Švéda, M. (2011). Practical IPv6 Monitoring Challenges and Techniques. In *Symposium on Integrated Network Management*, pages 660–663, Dublin, Ireland. IEEE CS.

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

Huang, D.-J., Yang, K.-T., Ni, C.-C., Teng, W.-C., Hsiang, T.-R., and Lee, Y.-J. (2012). Clock skew based client device identification in cloud environments. In *Advanced Information Networking and Applications*, pages 526–533.

Jøsang, A., Fabre, J., Hay, B., Dalziel, J., and Pope, S. (2005). Trust requirements in identity management. In *Proceedings of the 2005 Australasian Workshop on Grid Computing and e-Research Volume 44*, ACSW Frontiers '05, pages 99–108, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

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

Kumpošt, M. (2008). *Context Information and User Profiling*. PhD thesis, Masaryk University, CZ.

Laiping Zhao, Yizhi Ren, Mingchu Li, and Kouichi Sakurai (2012). Flexible service selection with user-specific QoS support in service-oriented architecture. *Journal of Network and Computer Applications*, 35(3):962–973. Special Issue on Trusted Computing and Communications.

Lanze, F., Panchenko, A., Braatz, B., and Zinnen, A. (2012). Clock skew based remote device fingerprinting demystified. In *Global Communications Conference*, pages 813–819.

Love, R. (2005). Kernel korner: Intro to inotify. *Linux Journal*, 2005(139).

McKeown, N., Anderson, T., Balakrishnan, H., Parulkar, G., Peterson, L., Rexford, J., Shenker, S., and Turner, J. (2008). OpenFlow: enabling innovation in campus networks. *SIGCOMM Computer Communication Review*, 38(2):69–74.

Megyesi, P. and Molnár, S. (2013). Analysis of elephant users in broadband network traffic. In Bauschert, T., editor, *Advances in Communication Networking*, Lecture Notes in Computer Science, pages 37–45. Springer Berlin Heidelberg. LNCS 8115.

Meints, M. and Gasson, M. (2009). *High-Tech ID and Emerging Technologies*, pages 130–189. Springer Berlin Heidelberg.

Murdoch, S. J. (2006). Hot or not: Revealing hidden services by their clock skew. In *Computer and Communications Security*, pages 27–36, New York, NY, USA. ACM.

Napatech (2014a). Time to ReThink Mobile Network Analysis. White paper, version 6, available online at http://www.napatech.com/sites/default/files/ dn-0720_ttrt_mobile_network_analysis_v06_us_a4_ online.pdf.

Napatech (2014b). Time to ReThink Performance Monitoring. White paper, version 6, available online at http://www.napatech.com/sites/default/files/dn-0645_ ttrt_performance_monitoring_ v6_us_a4_online.pdf.

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

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. and Franková, B. (2014). On reliability of clockskew-based remote computer identification. In *International Conference on Security and Cryptography*. SciTePress Science and Technology Publications.

Polčák, L. and Holkovič, M. (2013). Behaviour of various operating systems during SLAAC, DAD, and ND. http://6lab.cz/?p=1691.

Polčák, L., Holkovič, M., and Matoušek, P. (2013a). A New Approach for Detection of Host Identity in IPv6 Networks. *In Data Communication Networking*, pages 57–63. SciTePress Science and Technology Publications.

Polčák, L., Holkovič, M., and Matoušek, P. (Accepted, 2013b). Host Identity Detection in IPv6 Networks. In *Communications in Computer and Information Science*. Springer Berlin Heidelberg, DE.

Polčák, L., Jirásek, J., and Matoušek, P. (2013c). Comments on ”Remote physical device fingerprinting”. IEEE Transactions on Dependable and Secure Computing. Accepted, PrePrint available.

Sanguanpong, S. and Koht-Arsa, K. (2013). A design and implementation of dual-stack aware authentication system for enterprise captive portal. In *9th International Conference on Network and Service Management (CNSM)*, pages 118–121, Zürich, Switzerland.

Scarfone, K. A., Grance, T., and Masone, K. (2008). Computer security incident handling guide. Technical Report SP 800-61 Rev. 1., National Institute of Standards & Technology, Gaithersburg, MD, United States.

Sharma, S., Hussain, A., and Saran, H. (2012). Experience with heterogenous clock-skew based device fingerprinting. In *Workshop on Learning from Authoritative Security Experiment Results*, pages 9–18. ACM.

Thomson, S., Narten, T., and Jinmei, T. (2007). RFC 4862 *IPv6 Stateless Address Autoconfiguration*. Internet Engineering Task Force.

Wing, D. and Yourtchenko, A. (2012). *Happy Eyeballs: Success with Dual-Stack Hosts*. IETF. RFC 6555 (Proposed Standard).

Zander, S. and Murdoch, S. J. (2008). An improved clockskew measurement technique for revealing hidden services. In *Proceedings of the 17th Conference on Security Symposium*, pages 211–225, Berkeley, CA, USA. USENIX Association.