THE JOURNAL OF DIGITAL FORENSICS, SECURITY AND LAW

Journal of Digital Forensics, Security and Law

Volume 9 | Number 2 Article 2

2014

On Identities in Modern Networks

Libor Polcak
Brno University of Technology

Radek Hranick
Brno University of Technology

Tomas Martınek
Brno University of Technology

Recommended Citation

Polcak, Libor; Hranick, Radek; and Martınek, Tomas (2014) "On Identities in Modern Networks," *Journal of Digital Forensics, Security and Law*: Vol. 9 : No. 2 , Article 2.
DOI: https://doi.org/10.15394/jdfsl.2014.1167
Available at: https://commons.erau.edu/jdfsl/vol9/iss2/2

ON IDENTITIES IN MODERN NETWORKS

Libor Polčák, Radek Hranický, and Tomáš Martínek
Faculty of Information Technology, Brno University of Technology
Božetěchova 2, 612 66 Brno, Czech Republic
{ipolcak, ihranicky, martinto}@fit.vutbr.cz

ABSTRACT

Communicating parties inside computer networks use different kind of identifiers. Some of these identifiers are stable, e.g., logins used to access a specific service, some are only temporary, e.g., dynamically assigned IP addresses. This paper tackles several challenges of lawful interception that emerged in modern networks. The main contribution is the graph model that links identities learnt from various sources distributed in a network. The inferred identities result into an interception of more detailed data in conformance with the issued court order. The approach deals with network address translation, short-lived identifiers and simultaneous usage of different identities. The approach was evaluated to be viable during real network testing based on various means to learn identities of users connected to a network.

Keywords: lawful interception, intercept related information, content of communication, user identities, identification, linkability.

1. INTRODUCTION

Besides peaceful activities, computer networks are used for illegal actions or for communication of law offenders. *Lawful interception* (LI) (ATIS/TIA, 2006; ETSI, 2001) aims at gathering evidence from computer network communication. A court order may impose network operators or service providers to intercept all communication of a suspect. The collected evidence is later analysed by a *Law Enforcement Agency* (LEA). The evidence has to be indisputable and complete so that it can be used in a court room.

European standards for LI (ETSI, 2006) define reference architecture of a LI System (LIS). The entity carrying an interception in its network can be ordered to pass meta data about the communication of the interception target, i.e., discovered information about their network identifiers in a form of *Intercept Related Information* (IRI). In addition, the entity can be obliged to provide a copy of the communication of a suspect – *Content of Communication* (CC). The reference architecture embodies *Internal Interception Function* (IIF) divided into IRI-IIF and CC-IIF. IRI-IIF monitors the network and creates IRIs; CC-IIF copies the flows of a suspect in the form of CCs.

This paper proposes a model based on the graph theory that addresses several challenges that LI faces in modern networks. Firstly, the imminent exhaustion of IPv4 address space stimulated the increased need for *network address translators* (NATs) that often translate IPv4 addresses of several computers to only one IPv4 address and, consequently, conserves the IPv4 address space. However, the translation hinders the identification.

As of writing of this paper, four of the five regional registrars (all except the African AFRINIC) are already in the state in which IPv4 addresses are allocated according to very strict policies. The replacement, IPv6, poses several new challenges for LI. Firstly, a computer can generate thousands of IPv6 addresses and use them for simultaneous communication

(Narten, Draves, & Krishnan, 2007). Secondly, IPv6-enabled computers are often dual stacked and can intermix both IPv4 and IPv6 communication even for a single session (Sanguanpong & Koht-Arsa, 2013; Wing & Yourtchenko, 2012).

The model described in this paper handles both NAT and IPv6. More importantly, it can link all IPv4 and IPv6 identities and provide complete data in compliance with the court order.

The trend of last years is to bring application layer awareness into LI (AQSACOM, 2012; Hoffman & Terplan, 2006; Utimaco Safeware AG, 2010; Yang & Liu, 2013). Although the predominant protocols are related to e-mail and *Voice over IP* (VoIP), other protocols are also considered. Nevertheless, ETSI defines only application layer LI for e-mail (ETSI, 2010a) and multimedia services including VoIP (ETSI, 2010b). The model proposed in this paper is generic and it can be used with a range of application protocols.

Our solution deals with the challenges related to multiple identities (Pfitzmann & Hansen, 2010), determined by IP addresses, MAC addresses, application level identifiers, switch ports etc., by creating a graph of the network state. Discovered identifiers are stored as vertices. When two network identifiers are related, e.g., both belong to the same computer or person, there is an edge between the vertices in the graph. In addition, the vertices are categorised according to their relevance into several types, and consequently, our solution may link the identities and intercept all data allowed by a court order depending on its wording. The graph is built based on information learnt by modules scattered in the network. The modules discover identities from various sources, including network traffic and a cooperation with third party programs.

This paper is structured as follows. Section 2 lists the challenges in modern LI and explains how the proposed LI architecture addresses the challenges. Section 3 expands the architecture of a LIS and discusses possible sources of information in the network. Section 4 provides a formal model for inferring identities from various sources. We created a proof-of-concept LIS based on the guidelines and evaluated it in Section 5. Related work is discussed in Section 6. Section 7 proposes future work and Section 8 concludes the paper.

2. CHALLENGES

This section discusses the challenges of LI within modern networks. Additionally, it explains the contribution of this paper, i.e., how it helps in solving these challenges.

2.1 Challenges in IPv4 Networks

The shortage of IPv4 addresses starts to be visible in modern networks. In the past, home networks were usually assigned one public IPv4 address. Today, *carrier grade NAT* (CGN) or *multi-layer NAT* depicted in Figure 1 are becoming common. In such networks, home networks cannot be identified by a public IPv4 address. Instead, each flow has to be handled separately as flows originating from a single local address can be translated to more public IP addresses and several local IP addresses usually share a pool of public IP addresses. The proposed model contains extensions (Section 4.5) for networks with NAT, which handle all remote communication in NATed networks on the flow level.

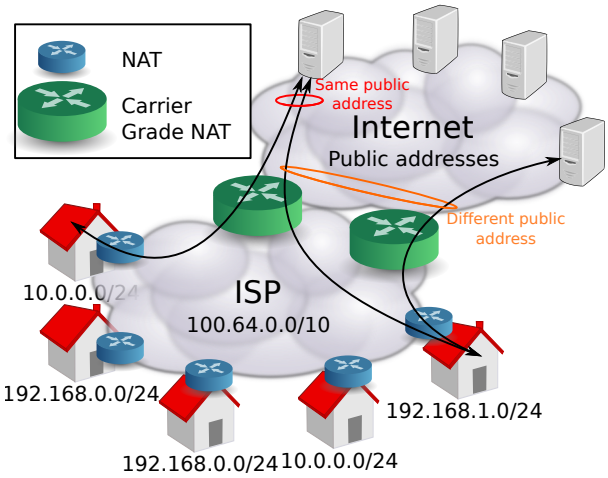

Figure 1 Multiple layers of NAT are becoming frequent in modern networks.

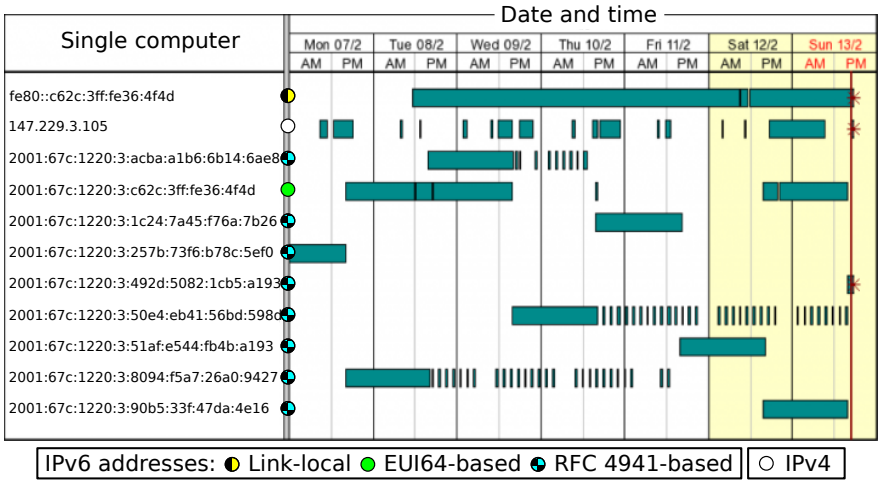

Figure 2 Multiple IPv6 addresses were used during a week by a single Windows-based computer simultaneously. The computer was not shut down during that week. While it used only one IPv4 address, it generated a new IPv6 address every day. The bars show the time periods when the addresses were actively used in the network.

2.2 Challenges in IPv6 Networks

IPv6 introduced (Narten et al., 2007) short-lived temporary addresses that can be generated by any computer in an 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.

Usually, even in IPv6-enabled networks, IPv4 is still present (as also illustrated in Figure 2). Recent operating systems and web browsers employ a mechanism called *Happy eyeballs* (HE) (Wing & Yourtchenko, 2012). HE keeps a record of the responsiveness of internet servers via IPv4 and IPv6. Based on the record, HE dynamically shifts between IPv4 and IPv6 and selects 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.

The proposed LIS deals with multiple identities in two ways. Firstly, we evaluated (Polčák, Holkovič, & Matoušek, 2013) several methods for learning the IPv6 addresses of one computer; including the detection of short-lived addresses. Secondly, the proposed graph model in Section 4.2 does not impose any limits on the number of IP addresses known for a single computer. Moreover, the proposed algorithms are aware of the possibility of a computer having multiple addresses learnt from different sources. Consequently, the proposed LIS can intercept data of all identities of one computer.

Overlay networks are built on top of the standard networks. Transition mechanisms (Carpenter, 2001; Despres, 2010; Huitema, 2006) aim at connecting IPv6 islands to the main IPv6 Internet through IPv4-only network. Consider the user in the IPv4 network in the Figure 3. The end host encapsulates IPv6 datagrams into IPv4 datagrams (depending on the

method either directly or using more protocols, e.g., UDP). The IPv6 datagrams are decapsulated by a specific node that provides the transition between IPv4 and IPv6 networks. The observed IP addresses can identify middle boxes instead of the real communicating party.

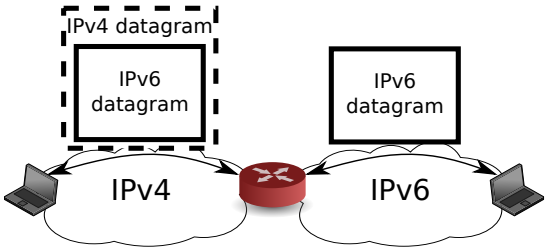

Figure 3 Tunneling is often used to extend IPv6 networks through IPv4-only networks.

2.3 Other Challenges

There are several kinds of court-approved interceptions. Sometimes it is allowed to intercept only data of a specific IP address. On other occasions, all traffic of a specific computer or of a specific user has to be intercepted. Some court orders allow to link identities whereas other court orders do not. The Scopes established in Section 4.1 address the issue of the wording of a court order.

An interception can target a specific application layer protocol, e.g., SIP (voice), e-mail or HTTP (web access). The proposed system is distributed and supports different sources of identities. Additionally, our research group develops FPGA-based probes. The goal is to integrate application level parsing into a high speed FPGA-based probe. However, the description of the probes is out-of-scope of this paper. Section 3 lists possible IAPs for deployment of LI probes in a network.

Cronin et al. (Cronin, Sherr, & Blaze, 2008) reported the problems related to packet-switched networks. They discussed *confusion* and *evasion of detection*. Confusion occurs when a surveillance system detects data transmissions but it is deceived into decoding false information. Evasion of detection happens when a surveillance oversights the communication. The proposed distribution of *intercept access points* (IAPs), i.e., the points in the network where the traffic is analysed or captured, addresses both confusion and evasion. The details are explained in Section 3.

Recently, new paradigms to control network, such as *Software Defined Networking* (SDN) (McKeown et al., 2008) and various overlay networks emerged. The modular approach to identity detection allows various sources of information. As a proof-of-concept, the LIS evaluated in Section 5 gathers information about the network from an SDN controller.

High mobility of users and recent growth of mobile devices connected to the Internet brings yet another challenge to LI. While we do not address this issue in detail in the paper, the modular input for identities can be in principle used in sharing information about identities of mobile users across several networks.

Last but not least, encryption is a major challenge for LI. Encrypted flows cannot be parsed on their path from source to destination without the knowledge of keys or a shared secret. Since the possibility of having scattered sources of information about identities is covered in the architecture described in Section 3, the architecture can deal with encryption in case the identities are learnt in cooperation with the accessed services.

3. LEARNING IDENTITIES

This section describes the proposed approach that deals with the challenges in LI performed in modern networks listed in Section 2. The described model enhances the reference architecture (ETSI, 2006), mostly IRI-IIF and partially CC-IIF.

The focus of IRI-IIF is on meta data while CC-IIF copies the traffic. Therefore, IRI-IIF needs to be optimized for parsing network traffic and dealing with specific protocols while CC-IIF needs to mirror the traffic to the *Mediation Function* (ETSI, 2006) data store.

We augment a detailed architecture of IRI-IIF. The architecture comprises of several parts:

- *IRI-Core* holds a centralised view on the

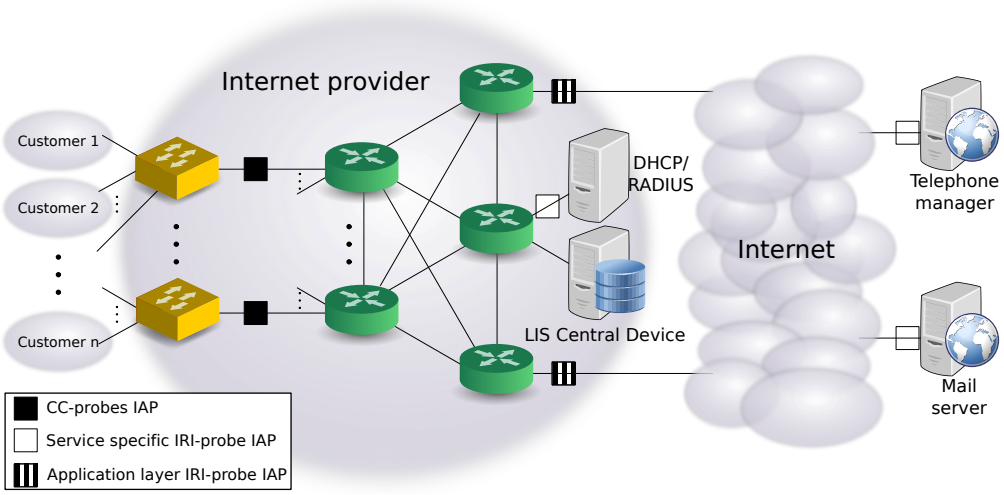

Figure 4 An example of a deployment of different Intercept Access Points (IAPs) in the network of an Internet access provider.

network. The proposed graph model is constructed by IRI-Core.

- *Application layer modules* parse all network traffic (they see) and pass discovered network identifiers and relations between them to IRI-Core.

- *Specific service modules* are distributed to specific positions in the network. Each module discovers mappings between network identifiers from a single protocol, e.g., DHCP, RADIUS, IPv6 neighbor discovery. The information is passed to IRI-Core. Both internal (module integrated to the service process) and external (passive monitoring of a mirrored or tapped traffic) modules are possible.

As CC-IIF needs to select the traffic for interception on wire speed, a CC-IIF probe has to identify the traffic quickly. Since network layer identifiers are preserved across the path (except NAT) and flows can be identified by an IP address pair and port number pair, the proposed method use these two types of identifiers to select the traffic to be intercepted by CC-IIF. Another benefit is that these identifiers are broadly supported by network gear manufacturers (e.g., access control lists supported by Cisco, Juniper, HP, etc.) and protocols for network control (e.g., OpenFlow).

Figure 4 shows an example network topology of an Internet provider. Its customers are connected to the core network. The core network is connected to the Internet with multiple edge routers.

A specific service module of IRI-IIF should be located near its corresponding service server if it is possible (e.g., DHCP, RADIUS) or distributed in the core network in relevant locations, e.g., IPv6 neighbor discovery (Polčák et al., 2013).

As the services (VoIP, e-mail, etc.) are usually located outside of the network, application layer modules monitor *Intercept Access Points* (IAPs) at the edge of the network. However, if these services were located in the Internet provider's network, a specific service module can be deployed on the link to the service or integrated on the server (similarly to the IAP near the RADIUS/DHCP server). IRI-Core is a part of the central device.

The idea behind the location of IRI-IIF IAPs is to move them close to the provided service to avoid traffic confusion (Cronin et al., 2008). Ideally, IRI-IIF IAPs are integrated within the service process. In this case, the identities seen by the LIS are guaranteed to be the same as

those registered with the service. Moreover, traffic encryption is also not a problem since the service process decrypts the traffic.

However, the integration often cannot be achieved in practice. Two reasons prevail. 1) The service is located outside of the jurisdiction of the LEA interested in the LI. 2) The service provider is concerned about the stability and performance of LI plug-ins. The IRI-IIF IAPs depicted in Figure 4 on the edge of the Internet provider network address the former issue. The IRI-IIF IAPs located near the service deals with the latter.

The IAPs for CC data needs to be located as close to the interception target as possible (Cronin et al., 2008). With CC-IIF IAPs in the access part of the network, it is possible to intercept even data that are exchanged between two customers, e.g., RTP part of a SIP call carried between the callers in the peer-to-peer manner. In addition, the interception is not limited to data of a specific service but broader scope of interception can be triggered when the identity of a target is discovered. See Section 4 for details.

4. TARGETING THE INTERCEPTION

An LI has to comply with law regulations and a specific court order. Whereas the regulations are quite stable, the wording of a court order may differ on a case-by-case basis. Sometimes a LEA knows a network identifier used by the interception target in the past (e-mail address, IP address, etc.) and it is interested in all communication of the same person. On other occasions, LEAs are interested only in the communication identified by a specific identifier. This section deals with the issues arising from the wording of court orders of interceptions and introduces a graph-based view on the state of the network and algorithms to select the network identifiers for an interception.

4.1 Interception Scope

From the interviews with representatives of LEAs in our country, we have identified three Scopes of interception:

1= *Specific identifier*: intercepts just the communication directly identified by the given identifier.

2= *Specific computer*: intercepts all communication of a computer that is related to the given identity. For example, this Scope intercepts IPv6 traffic of a computer identified by an IPv4 address or it intercepts all traffic of a computer when a given e-mail address was used by that computer.

3= *Specific user*: intercepts all traffic of a specific user, e.g., when his or her identity is disclosed by an application layer identifier, or, when he or she authenticates with RADIUS or other authentication technique.

The defined Scopes treats network identifiers in a different manner depending on their meaning. We identified five Categories of network identifiers:

α= Flows defined by a 5-tuple consisting of:

- local and remote IP address,
- transport level protocol identifier, e.g., TCP or UDP,
- local and remote port number.

β= Network layer addresses, e.g., IPv4 or IPv6 addresses.

γ= Physical layer interface addresses, e.g., MAC addresses; and any other identifiers specifying a unique computer, e.g., DHCPv6 DUID.

δ= Authentication user names (RADIUS, PPP, etc.).

λ= Application level identifiers (login, SIP id, e-mail address, etc.).

Let us define the set of Categories, $C = \{\alpha, \beta, \gamma, \delta, \lambda\}$, and the partial ordering $\delta > \gamma > \beta > \alpha$ and $\lambda > \alpha$; meaning α has the narrowest coverage whereas δ and λ represent the broadest

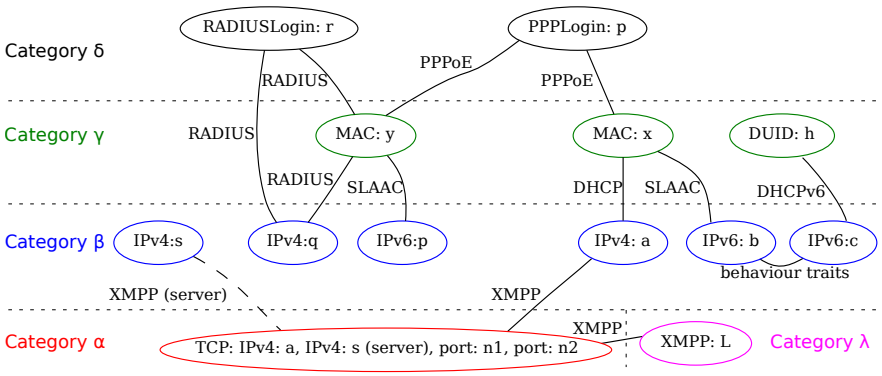

Figure 5 An example graph of a network state.

coverage. For example, a user can authenticate a few devices, each of them having a specific IP address (Category β identifier), using only one token (Category δ identifier). Another example is an interface with one physical address (Category γ) initiating connections using more than one IP address (as discussed in Section 2). The final example is a user signed to a service using a login of Category λ on a private phone and a shared computer. Each device carries the communication in separate Category α flows.

4.2 Graph Interception Model

To enable interceptions of the above defined Scopes, we propose to manage a graph model of the identities known in a specific state of the network. The identities are identified (Pfitzmann & Hansen, 2010) by identifiers of the specified Categories. The graph model of the state is defined by the following undirected graph $S = (V, E, p, l)$ where:

- V is a set of vertices. Every vertex represents a network identifier.
- $E \subseteq V \times V$ is the symmetric adjacency relation between the vertices.
- $p : E \to P$ is a total function that maps each edge to a protocol from a set of all supported protocols P. For example, $p((K.L.M.N, x@y))$ yields SMTP if the relation of an IP address $K.L.M.N$ and an e-mail address $x@y$ was learnt from SMTP.
- $l : V \to C$ is a total function that maps each vertex to its Category.

When a LIS has to determine CC-IIF identifiers (IP addresses or flow identifiers) for a given interception target identified by a network identifier i, there has to be a node $v \in V$ in the state graph, such as v represents i. If there is no such v, i is not known to be present in the network and there are no data to be intercepted. The following text treats terms network identifier and vertex interchangeably.

4.3 Network State Example

Let us consider the graph depicted in Figure 5. A user with a PPP login p connected two computers to the network; one of the computers is authenticated with a RADIUS server. Both computers have several IP addresses obtained from different sources (DHCP, DHCPv6, RADIUS) or automatically generated by the computers (SLAAC). DHCPv6 uses a special identifier called DUID that does not have to be related to the MAC address of the computer. Nevertheless, it is possible to detect the relation between IPv6 addresses b and c from other sources, e.g., behavioural analysis, switch port etc. Additionally, the user opened an XMPP connection to a server s using the IP address a. Note that the relation between the Category α identifier and IP address s is dashed, meaning that it is not actually a part of E, and consequently, it is not used to link identities.

Let us demonstrate the scope of interceptions targeted on XMPP login L (Scopes 1, 2, and 3)

and IPv4 address q (Scopes 1' and 2') in Figure 6.

Scope 1 interceptions gather only data identified by the target identifier and related identifier in its coverage. Therefore, the interception aimed at the XMPP login L covers the flow related to the L but other traffic of the IP address a is not intercepted. Correspondingly, the other interception of q gathers packets containing the IP address q, other packets of these computers are ignored.

In contrast, the Scope 2 interceptions cover all identifiers of a specific computer. Specifically, the interception targeted on the login L covers all identities related to and only to the computer in the right, including IP addresses a, b, and c and the XMPP flow. Similarly, the interception targeted on IPv4 address q includes the other IP address p of the left computer but the right computer does not belong to the Scope.

Scope 3 aims on specific users. The intercept targeting XMPP login L aims only on the flows that are a part of a session related to L. As the IP address of q does not represent an identity of a user, Scope 3' interception is not defined.

4.4 Definitions of Graph Operations

When a vertex v exists for an identifier i that should be intercepted, it is necessary to determine all identifiers that are relevant to the intercept according to the specified Scopes 1–3. Let us define a function capture : $V \to 2^V$ that yields the relevant identifiers for a capture targeted at any given $v \in V$.

Firstly, let us define binary relations used to construct capture. The Formula 1 defines the antisymmetric binary relation covers $\subseteq V \times V$. Vertices x and y are related in case they are directly connected and y is of a narrower Category.

The Formula 2 defines the binary relations $\text{linked}_c \subseteq V \times V$, one for each $c \in C$. A vertex x is in relation linked_c with a vertex y if they are directly connected by an edge and vertex y is of the Category c or narrower. Note that the network state cannot contain edges between an IP address of a server and a flow of the Category α so that it is not possible to cross to a domain of another computer. If this condition holds, the reflexive and transitive closure linked_c^* does not link the identities of other communicating parties through flows and servers.

Finally, let us define capture in Formula 3 using the reflexive and transitive closure of the relations covers and linked_c for $c \in \{\gamma, \delta, \lambda\}$.

As stated in Section 3, the proposed LIS architecture aims only at intercepting data according to identifiers from Category α or β. Therefore the final capture_{CC} function is defined in Formula 4. However, the results of capture are more generic since they can be used even without these limitations.

4.5 Considerations for Network Address Translators

The described model is suitable for networks without network address translation (NAT). However, when NAT is located in between the IRI-IIF IAP and the CC-IIF IAP, several issues has to be considered. Fortunately, the issues can be addressed with specific rules during a network state graph construction:

1. The graph cannot contain edges between the public nodes of Category β, γ, δ or λ and a node of Category α, e.g., a public IP address or application login and the corresponding flow (similarly to edges between a flow and its server address).

2. In contrast, following edges have to be added when a local flow a is translated to a public flow b:

   (a) an edge between flow a and b;

   (b) an edge between flow a and:

   - related local identifiers of Categories β and γ (public identifiers belongs to the translator),
   - all related identifiers of Categories δ and λ.

3. There is no edge between local flow a and the remote server IP address.

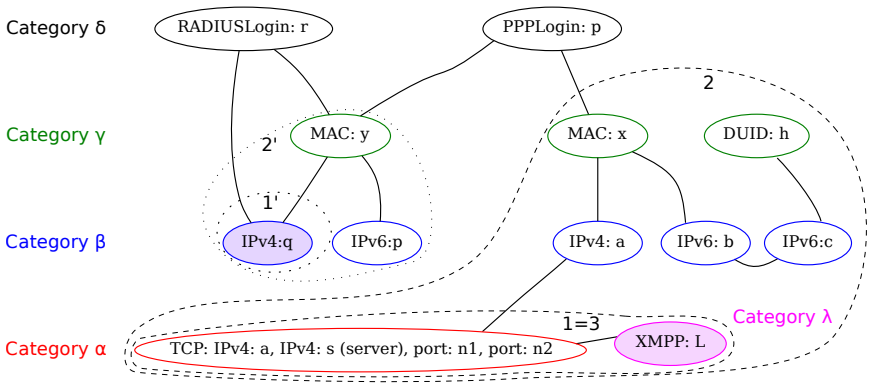

Figure 6 An example of intercepts in the network state.

$$\text{covers} = \{(x, y) \in V^2 : (x, y) \in E \wedge l(x) > l(y)\}. \tag{1}$$

$$\text{linked}_c = \{(x, y) \in V^2 : (x, y) \in E \wedge l(y) \leq c\}. \tag{2}$$

$$\text{capture}(v) = \begin{cases} \{x \in V : v \,\text{covers}^*\, x\} & : \text{Scope 1 (specific identifier),} \\ \{x \in V : v \,\text{linked}^*_\gamma\, x\} & : \text{Scope 2 (specific computer),} \\ \{x \in V : v \,\text{linked}^*_{l(v)}\, x\} & : \text{Scope 3 (specific user) and } l(v) \in \{\delta, \lambda\}. \end{cases} \tag{3}$$

$$\text{capture}_{CC}(v) = \{x \in V : x \in \text{capture}(v) \wedge l(x) \in \{\alpha, \beta\}\}. \tag{4}$$

$$\text{capture}(v) = \{x \in V : v(\text{covers} \cup \text{linked}_\alpha)^* x\} : \text{Scope 1 in the NAT scenario.} \tag{5}$$

Since these rules insert edges connecting local and remote flows, a slightly updated version of capture has to be applied in the Scope 1 case (for other Scopes, Formula 3 still applies). The updated version is defined by the Formula 5.

Figure 7 illustrates the above mentioned rules on an example of a SIP call. A user identified with a private IPv4 address *local* is registered with a remote server s and the user is calling another user on IP address r. These two flows are identified by NAT. Since the dashed edges in Figure 7 are not considered by the function capture (defined in formulae 3 and 5), the computer shown in Figure 7 cannot be linked with another computer that happens to have its IP address translated to the same public address.

The function capture yields the correct results according to the specified Scope of the interception. For example, the interception can be targeted on the access login to the network (PPPLogin p). Flows with both local and public IP address of the target user machine are returned by the function capture while only the local address Category β identifier is returned. The returned vertices are the same for all Scopes; they are shown with the colored background.

The benefit is that the identifiers yielded by capture can be used by a CC-IIF probe in any location in the network, inside or outside of the NATed network. In the example in Figure 7, a probe outside of the NATed network captures

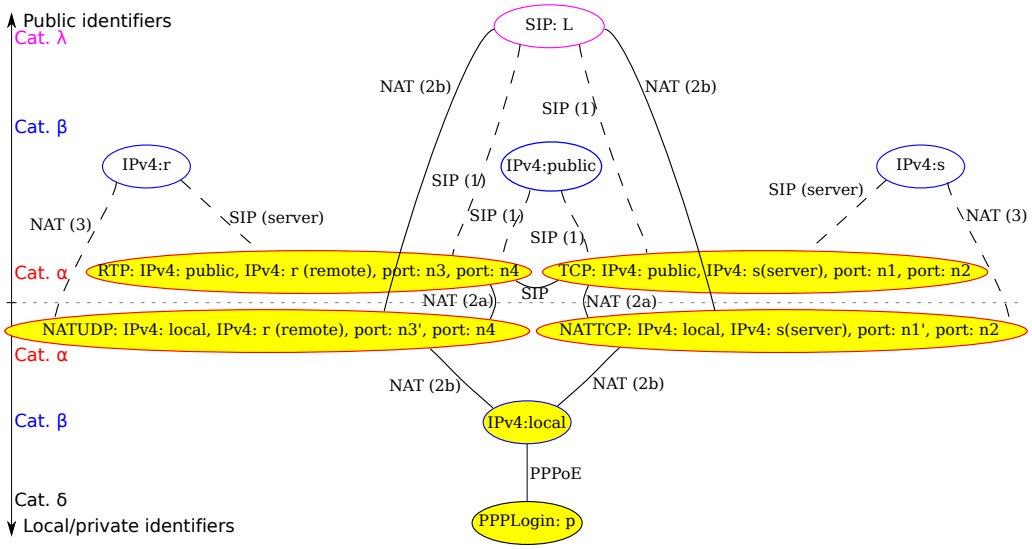

Figure 7 An example of a network state with NAT. The rules for a presence or an omission of an edge are referred.

according to the public identifiers (flows in the upper part of the Figure 7). Such a probe does not see any traffic identified by the private network identifiers depicted in the lower part of the Figure 7. On the other hand a probe located inside of the NATed network sees just the identifiers displayed in the bottom part of the Figure 7.

Note that the specific service module for NAT has to advertise the network translation on the flow level.

5. PROOF-OF-CONCEPT

To evaluate the proposed architecture, we implemented a LIS following the approach introduced in this paper. *Intercept Related Information – Internal Interception Function* (IRI-IIF) (ETSI, 2006) in the LIS is highly modular and modules for specific protocols can be distributed across the network. The discovered identities of various protocols can be linked and the Scope of an intercept can be defined according to the wording of the court order for the intercept. All three interception Scopes introduced in Section 4 are supported.

As the number of protocols used in networking is high, we had to limit the number of protocols that we evaluated in our proof-of-concept LIS. The goal was to show that the proposed solution is viable for:

- protocols for authentication (RADIUS);
- protocols used to access the Internet from households (PPPoE used in ADSL);
- linkability of IPv4 and IPv6 addresses of the same computer, addresses are learnt from DHCP, DHCPv6, neighbor discovery (Polčák et al., 2013), and OpenDaylight controller;
- application layer protocols (instant messaging, e-mail etc.);
- cooperation with third party programs (in our case, we include OpenDaylight SDN controller as a source of information about computers in the network: IP addresses and location of computers in the network).

The Figure 8 shows the various modules that were deployed in our laboratory environment and University networks during our evaluations. All modules passed the detected identifiers to the central device where they were linked and the graph of the current state of the network was constructed.

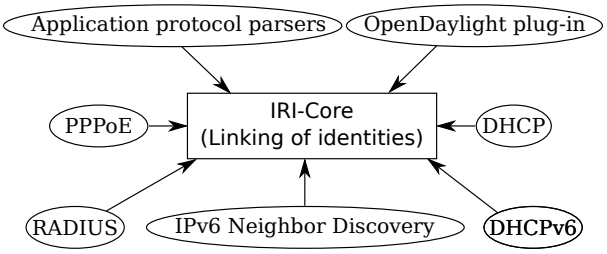

Figure 8 The ETSI IRI-IIF architecture deployed during the testing.

The Table 1 shows the identifiers and their Categories that we evaluated.

Table 1 Network identifiers supported by the proof-of-concept LIS.

Category	Identifiers
α	TCP flow
β	IPv4 and IPv6 address
γ	MAC address, DHCPv6 DUID
δ	PPP login and session; RADIUS login
λ	e-mail address; OSCAR, YMSG, and XMPP login; IRC login and channel

During the testing, we connected devices under our control and checked that the assigned identifiers were registered by IRI-IIF located at the central device. Additionally, we checked that the detected identifiers were correctly linked, including the IP addresses simultaneously used by a single computer. Later, we initiated several application layer sessions and again checked that the flows were correctly represented in the graph. The testing was successful as we were able to link the identities of our computers and testing users gathered from various sources.

Finally, we configured our 1 and 10 Gbps probes distributed in the testing network. As all traffic of the computers under test flew through the probes, the probes captured all packets related to the configured interceptions.

However, the quality of interceptions in real-life scenario depends on several factors:

- The application layer and specific service modules has to provide accurate data to the IRI-Core. When some identities are not known inside IRI-Core, data related to these identities might evade the interception. In contrast, incorrect detection of identities might link unrelated identities and consequently, packets related to the incorrectly linked identities might be captured as a part of one interception.
- CC-probes has to see all traffic that can be a part of an interception. Moreover, the probes has to capture data on wire-speed.

In addition, the definition of capture for level 3 interceptions (in the Formula 3) aimed on Category λ identifiers yields only the identities directly linkable to the input Category λ identifier. As a consequence, data produced by the same user in a different application are not intercepted (unless there is a module that links the identifiers in the graph). Depending on the aim of the interception, this might be in conformance with the wording of the interception, or, data produced by the other applications might be incorrectly missing.

6. RELATED WORK

This work is primarily based on ETSI standards for lawful interception (LI) (ETSI, 2001, 2006, 2010a, 2010b). These standards specify a generic architecture, which is in nature very similar to U.S. ATIS/TIA (2006) standards for CALEA-based LI. We detailed the architecture of *Intercept Related Information – Internal Interception Function* (IRI-IIF) (ETSI, 2006) that detects identities in the network.

The proposed approach of learning the identities in a distributed manner is compatible with other solutions with roots in ETSI or ATIS/TIA standards, such as the architecture of Cisco Systems (Baker, Foster, & Sharp, 2004), Aqsacom ALIS (AQSACOM, 2012) or UTIMACO LIMS (Utimaco Safeware AG, 2010). In comparison to these works, we list more challenges, namely related to IPv6. Additionally, we propose solutions for their concerns, especially those related to target identification.

Other previously proposed monitoring architectures for LI (Karpagavinayagam, State, & Festor, 2007; Milanović et al., 2003; Yang & Liu, 2013) were tailored for specific protocols, mostly voice related. The architecture used in this paper is generic and it is suitable for identifiers from all networking layers.

This work addresses the concerns of Cronin et al. (2008) related to confusion and detection of evasion from LI by locating the identity detection probes as close to the accessed service as possible while CC-probes are located close to the target.

7. FUTURE WORK

Currently, our research group is working on several extensions for the described LIS. Firstly, high-speed networks require hardware-based solutions for LI. The distributed structure of IRI-IIF was designed with the expectations of using IRI-probes to gather information about the identities in the network directly from the network traffic. We are currently working on high-speed FPGA-based 10–100 Gbps probes.

In addition, our research group is working on the detection of identities in a network with NAT. This will provide data for the mechanism described in Section 4.5. Furthermore, we are also working on improvements in local (Polčák et al., 2013) and remote (Polčák & Franková, 2014) detection of all addresses used by one computer. The remote detection includes unique traits of specific computers.

Furthermore, we are investigating the advantages that SDN brings to LI. Not only can SDN-switches be used as an IAP for CC, but the knowledge of network topology can also be utilised for lawful interception, e.g., to adjust the configuration of specific probes in the network based on the up-to-date topology.

Finally, our colleagues are working on tools for decoding of intercepted data.

8. SUMMARY

LI faces several challenges in modern networks. The shortage of IPv4 addresses results in extensive use of network address translation and the arrival of IPv6, a successor of IPv4. Consequently, a single computer simultaneously uses more than one IP address for its communication. In addition, interception of application protocols is becoming common. However, the services are often provided abroad, outside of the jurisdiction of a specific LEA. Furthermore, court-approved interceptions use specific wording for each case.

In this paper, we propose a distributed LIS that gathers information from various sources. The identities can be linked according to the presented algorithms based on the graph theory. The proposed approach is suitable for modern networks as it tackles the challenges of network translation, temporary IPv6 addresses, and simultaneous communication of one computer identified by different IP addresses. The proposed deployment of network probes aims to provide as complete data as possible even if the target accesses a service located outside of the jurisdiction of a LEA. The linking of identities can be used to obtain more data in conformance with the wording of a specific intercept court order.

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. We would like to thank all colleagues from our research group for both their valuable input on the architecture and identity handling described in this paper and their contribution during the development of the service specific modules. Last but not least, we would like to thank the reviewers, especially the one that indicated an inconsistency in the original definition of the partial ordering of the Categories C.

REFERENCES

AQSACOM. (2012). *Lawful Interception for IP Network.* (White Paper)

ATIS/TIA. (2006). *Lawfully Authorized Electronic Surveillance. J-STD-025-B.*

Baker, F., Foster, B., & Sharp, C. (2004). *Cisco Architecture for Lawful Intercept in IP Networks.* (RFC 3924)

Carpenter, B. E. (2001). *Connection of IPv6 Domains via IPv4 Clouds.* (RFC 3056)

Cronin, E., Sherr, M., & Blaze, M. (2008). On the (un)reliability of eavesdropping. *International Journal of Secure Networking*, *3*, 103–113.

Despres, R. (2010). *IPv6 Rapid Deployment on IPv4 Infrastructures (6rd).* (RFC 5569)

ETSI. (2001). *ETSI TR 101 943: Telecommunications security; Lawful Interception (LI); Concepts of Interception in a generic Network Architecture.* (Version 1.1.1)

ETSI. (2006). *ETSI TR 102 528: Lawful Interception (LI); Interception domain Architecture for IP networks.* (Version 1.1.1)

ETSI. (2010a). *ETSI TS 102 232-2: Lawful Interception (LI); Handover Interface and Service-Specific Details (SSD) for IP delivery; Part 2: Service-specific details for E-mail services.* (Version 2.5.1)

ETSI. (2010b). *ETSI TS 102-232-5: Lawful Interception (LI); Handover Interface and Service-Specific Details (SSD) for IP delivery; Part 5: Service-specific details for IP Multimedia Services.* (Version 2.5.1)

Hoffman, P., & Terplan, K. (2006). *Intelligence support systems: Technologies for lawful intercepts.* Auerbach Publications, U.S.

Huitema, C. (2006). *Teredo: Tunneling IPv6 over UDP through Network Address Translations (NATs).* (RFC 4380)

Karpagavinayagam, B., State, R., & Festor, O. (2007, June). Monitoring Architecture for Lawful Interception in VoIP Networks. In *Internet monitoring and protection.*

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

Milanović, A., Srbljić, S., Ražnjević, I., Sladden, D., Skrobo, D., & Matošević, I. (2003). Distributed system for lawful interception in VoIP networks. In *Eurocon 2003. computer as a tool.* (Vol. 1, pp. 203–207).

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

Pfitzmann, A., & Hansen, M. (2010). *A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management* (Tech. Rep.). Retrieved from `https://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf` (Version 0.34)

Polčák, L., & Franková, B. (2014). On reliability of clock-skew-based remote computer identification. In *11th international conference on security and cryptography.* Vienna, AT: SciTePress - Science and Technology Publications.

Polčák, L., Holkovič, M., & Matoušek, P. (2013). A New Approach for Detection of Host Identity in IPv6 Networks. In *Data communication networking* (pp. 57–63). Reykjavk, IS: SciTePress - Science and Technology Publications.

Sanguanpong, S., & 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* (pp. 118–121). Zürich, Switzerland.

Utimaco Safeware AG. (2010). *Lawful interception in the digital age: Vital elements of an effective solution.* (White Paper)

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

Yang, M., & Liu, H. (2013). Implementation and performance of VoIP interception based on SIP session border controller. *Telecommunication Systems*, 1–17.

