KUNGL TEKNISKA HÖGSKOLAN
Royal Institute of Technology

Privacy in the next generation Internet

Data protection in the context of European Union policy

Alberto Escudero-Pascual

Telecommunication System Laboratory
Department of Microelectronics and Information Technology
Royal Institute of Technology
Stockholm, SWEDEN

KUNGL TEKNISKA HÖGSKOLAN
Royal Institute of Technology

Privacy in the next generation Internet

Data protection in the context of European Union policy

Alberto Escudero-Pascual

A thesis submitted to
the Royal Institute of Technology
in partial fulfillment of the requirements for
the Doctorate of Technology degree.

December 2002

TRITA-IMIT-TSLAB AVH 02:01
ISSN 1651-4114
ISRN KTH/IMIT/TSLAB/AVH-02/01--SE

Department of Microelectronics and Information Technology
Telecommunication Systems Laboratory
Royal Institute of Technology
Stockholm, Sweden

Sometimes the first duty of intelligent men is the restatement of the obvious
Eric Arthur Blair

Abstract

With the growth in social, political and economic importance of the Internet, it has been recognized that the underlying technology of the next generation Internet must not only meet the many technical challenges but must also meet the social expectations of such a pervasive technology.

As evidence of the strategic importance of the development of the Internet, the European Union has adopted a communication to the Council and the European Parliament focusing on the next generation Internet and the priorities for action in migrating to the new Internet protocol IPv6 and also a new Directive (2002/58/EC) on 'processing of personal data and protection of privacy in the electronic communication sector'. The Data Protection Directive is part of a package of proposals for initiatives which will form the future regulatory framework for electronic communications networks and services. The new Directive aims to adapt and update the existing Data Protection Telecommunications Directive (97/66/EC) to take account of technological developments. However, it is not well understood how this policy and the underlying Internet technology can be brought into alignment.

This dissertation builds upon the results of my earlier licentiate thesis by identifying three specific, timely, and important privacy areas in the next generation Internet: unique identifiers and observability, privacy enhanced location based services, and legal aspects of data traffic.

Each of the three areas identified are explored in the eight published papers that form this dissertation. The papers present recommendations to technical standarization bodies and regulators concerning the next generation Internet so that this technology and its deployment can meet the specific legal obligations of the new European Union data protection directive.

In summary, the eight papers of this dissertation show:

- how eavesdroppers will be able to identify and track packets that belong to a particular node and the limitations of the privacy extension for stateless address autoconfiguration which in fact fails to provide privacy.
- a network architecture that provide unlinkability between a user's personal identifiable information and location information.
- a critical review of the policy initiatives to extend traditional powers of lawful access to communications traffic data and the European Union Data Protection Telecommunications Directive.

The dissertation concludes by presenting future work identified based on examining these three different areas.

Stockholm, 11th September 2002

Acknowledgments

I have left the acknowledgments section of this dissertation unwritten until the very last moment.

I have begun to suspect that my main fear is not to forget many of the people who supported me during these last three, very intensive, years of my life. My main fear is to be unfair.

I would be unfair if I didn't mention my family. They gave me the opportunity to view the world with critical eyes from the very beginning. I always found them supportive of my own decisions even though some of those decisions put us many kilometers away from each other.

I would be unfair if I didn't mention my research advisors. They not only gave me the resources, visions and guidance I needed but, in addition, they pushed me to discover and surpass my own limits.

I would be unfair if I didn't mention my colleagues in the Lab with whom I shared daily coffee breaks. They provided a much needed balance for me throughout endless days of work.

I would be unfair if I didn't mention all those who gave me the opportunity to visit the five continents to talk about my work. Every trip helped me to see global problems from a local perspective.

I would be unfair if I didn't mention my other big global family, the family that is always waiting for me in the world's many airports and train stations whenever I travel.

I would be unfair if I didn't mention all those close people that have had to cope with, if not suffer, my hectic life.

At the same time, it would not be fair to write a thesis ignoring the huge amount of economic, human and environmental resources that a mere hundred pages requires: more than 40 flights, over 5000 pages of paper and an unquantifiable amount of coffee.

This work is the result of a very long trip that has just started.

Alberto

Contents

1 Introduction

Although the Internet is rapidly becoming "the" communication network, it was not really engineered to preserve certain types of privacy. In keeping with the European Union policies regarding data protection there is a need to understand the benefits and to reduce the privacy risks of this new generation of Internet technology.

As evidence of the strategic economic and social importance of the development of the Internet, on the 21st of February 2002, the European Commission adopted a communication to the Council and the European Parliament, focusing on the next generation Internet and the priorities for action in migrating to the new Internet protocol IPv6 [1] .

Maintaining proper confidentiality with respect to location information, traffic information, and the actual data traffic itself are three of the key provisions of the new European regulatory framework for electronic communications infrastructure and associated services [2].

The European Union has just updated the Data Protection Directive to take into account new technological developments and empower users to take control of their personal identifiable information [3,4,5]. However, it is not well understood how this policy and the underlying Internet technology can be brought into alignment. For example, the current method in IPv6 of automatically configuring an Internet device [6] results in an identifier that is readily observable and recognizable - despite the user moving from one network to another. Privacy advocates have already pointed to this as a problem with respect to traffic analysis and location privacy [7,8,9] .

With the growth in social and economic importance of the Internet, it is recognized that the underlying technology of the next generation Internet must not only meet the many technical challenges (such as reliability or availability), but must also meet the social expectations of such a pervasive technology. These social expectations are now in the process of being embodied as regulations and law.

Thus although there have been many technical efforts to insure data confidentiality in the next generation Internet, it is still not known if the new IPv6 security and mobility features will actually be enough to empower users and protect their privacy or if in fact just the opposite will occur.

1.1 Criteria for improvement of IPv6 privacy

A definition of privacy introduced by Alan Westin [10] states:

"Privacy is the claim of individuals, groups and institutions to determine for themselves, when, how and to what extent information about them is communicated to others".

Trying to find a metric for the different privacy measures M. Köhntopp and A.

Pfitzman came up with a proposal for terminology [11] that defines: anonymity, unlinkability, unobservability, and pseudonymity. In general we assume that privacy solutions that work in the direction of unobservability are better than the ones that provide just anonymity or pseudonymity. When two solutions provide identical privacy attributes we consider a better solution the one that makes use of less computing and/or network resources. When talking about privacy in IPv6 networks we will consider a improvement in terms of privacy occurs when:

1. A solution (protocol or network architecture) empowers the user to determine which information is communicated to others. (opt-in versus opt-out)

2. A solution makes it less obvious (observable) that a given user wants to protect their privacy.

3. A solution reduces the ability of an eavesdropper to identify a specific user's traffic by observations at any point between the source and the destination.

4. A solution requires less network and computing resources to achieve the same degree of protection of personal identifiable information.

1.2 Privacy in the next generation Internet

This dissertation builds upon my previous licentiate thesis [12] concerning mobile privacy. This licentiate presented an extension to Zero Knowledge Systems' Freedom protocol [13,14], which provided seamless mobility and location privacy. The main goal of the protocol extension was to provide unlinkability between the mobile's identifiable information and the contents of the communication. By location privacy in the context of this dissertation we mean the capability of a mobile node to conceal the relation between location and personal identifiable information from third parties.

This dissertation is organized as a collection of published papers. In these papers I have concentrated on examining new emerging privacy challenges concerning the next generation Internet with the aim of providing recommendations to technical standarization bodies and regulators so that this technology and its deployment can meet the specific legal obligations of the new European Union data protection directive.

The first step in this direction was taken by identifying new privacy threats in IPv6 [Paper #2] and a revision of the proposed European Union Directive on 'processing of personal data and protection of privacy in the electronic communication sector' COM(2000)385 [3,4]. The goal was not only to identify the new possible emerging threats to privacy in IPv6, but also examine if the European Union's data protection legal provisions [2] (as part of the government's update of legislative telecommunication frameworks) are suitable to deal with new communications infrastructures.

Three timely and important privacy areas where identified during the initial legal and technical review:

1. **Unique Identifiers and observability:** The use of unique identifiers in telecommunication terminal equipment and the limitations of the different privacy extensions in IPv6.

2. **Privacy enhanced location based services:** Location privacy in Location Based Services and the role of mix networks in location privacy.

3. **Legal aspects of data traffic:** The legal treatment of data traffic and location data with respect to the European Union data protection policy.

These issues were explored in the papers described below.

1.3 Organization of the collection of papers

The dissertation is composed of 8 published papers:

1. A. Escudero, M. Hedenfalk, and P. Heselius, Location Privacy in Mobile Internet - An extension to Freedom Network. Internet Society Conference (INET2001). Stockholm, Sweden. June 2001.

2. A. Escudero, Location Privacy in IPv6: 'Tracking binding updates'. Tutorial at Interactive Distributed Multimedia Systems (IDMS2001). Lancaster, UK. September 2001.

3. A. Escudero, Requirements for unobservability of privacy extension in IPv6. Radio Vetenskap 2002. Stockholm, Sweden. June 2002, pp. 58.

4. A. Escudero, Privacy enhanced architecture for location based services in the next generation wireless networks. 11th IEEE Workshop on Local and Metropolitan Area Networks (LANMAN2002). Stockholm, Sweden. August 2002, pp. 169-172.

5. A. Escudero and G.Q. Maguire Jr., Role(s) of a proxy in location based services. 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications. (PIMRC2002). Lisbon. Portugal. September 2002, Vol.3 pp. 1252-1257.

6. A. Escudero and I. Hosein, The hazards of technology-neutral policy: questioning lawful access to traffic data. To appear in Communications of the Association for Computer Machinery (CACM) Journal. Accepted on the 5th September 2002 - Reviewed 19th October 2002.

7. A. Escudero, Privacy in mobile Internet in the context of the European Union data protection policy. Internet Society Conference (INET2002). Washington DC. USA. June 2002.

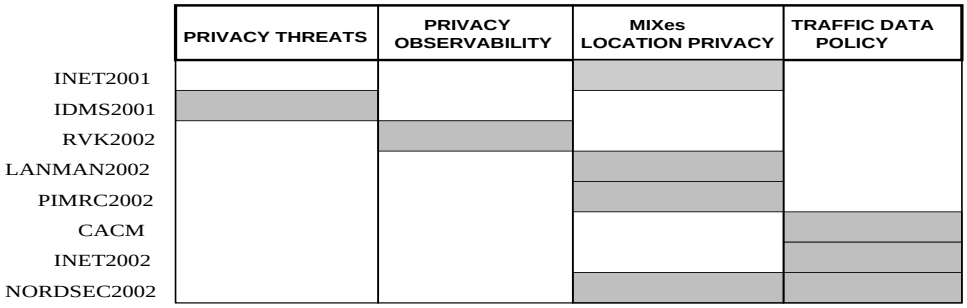

Figure 1: Papers and its areas

8. A. Escudero, T. Holleboom, and S. Fischer-Huebner, Privacy for location data in Mobile Networks (NORDSEC2002). Karlstad, Sweden. November 2002, pp. 220-232.

The papers are organized as follows [See Fig. 1]:
Paper #1 which was also part of my licentiate thesis provides the necessary background to this work in the area of mobility in mix networks. Paper #2 describes a set of privacy threats in the next generation Internet. Papers #3 to #7 focus on the three identified privacy areas: unique identifiers and observability (Paper #3), location based services and mix networks (Papers #4 and #5) and legal aspects of traffic data (Paper #6 and #7). Finally, the latest paper (Paper #8) summarizes most of the results in the different areas.

Note that I believe that the easiest way to read this dissertation is to start by reading Paper #8 as it summarizes most of the results and is my the latest work, then proceed with Papers #2 to #7.

2 Summary of original work

The objective of this section is to present a short summary of the appended publications forming this dissertation thesis and their novel contribution.

Paper #1

- **Location Privacy in Mobile Internetworking: Protocol extensions to Freedom Network.**
  Alberto Escudero, Martin Hedenfalk, and Per Heselius
  Internet Society's 11th Annual INET Conference (INET2001). A Net Odyssey - Mobility and the Internet, Stockholm, Sweden. June 2001.

The first paper was part of my licentiate thesis [12], the paper describes a set of protocol extensions to the Freedom System architecture to permit a mobile node to seamlessly roam among IP subnetworks and media types whilst remaining untraceable and pseudonymous.

The focus of this previous work was to try to prevent linkability between the location of wireless users and their activities in the Internet. *Flying Freedom* is a protocol extension to a pseudonymous IP network architecture called the Freedom System developed by the Canadian company Zero Knowledge Systems Inc. The Freedom System is a pseudonymous IP network that provides privacy protection by hiding the user's real IP addresses, email addresses, and other personal identifying information from *both* communication partners and eavesdroppers.

Our initial ideas were in the direction of integrating $MobileIP_{v4}$ into the Freedom System by encapsulating registration and deregistration messages and IPIP/GRE tunnels into Freedom Traffic [15,16]. Further studies and preliminary results showed that it was more adequate to extend Freedom to provide the same functionalities of $MobileIP_{v4}$ while providing the flexibility of rebuilding partial routes hiding the mobility associated with certain pseudonymous [Fig. 2].

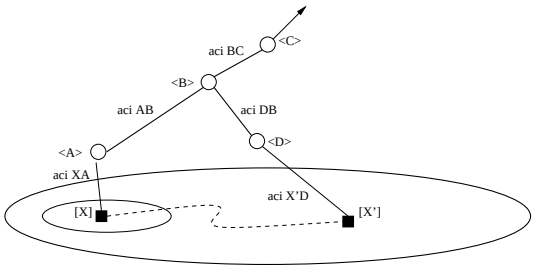

Figure 2: Mobility extensions for the Freedom System. The virtual circuit is partially recreated during a vertical handover $[X] \rightarrow [X']$. The exit node $< C >$ is not aware of any mobility.

In this paper we also introduce the possibility of having an *unlocated mobile server* roaming behind the Freedom System. The mobile server is able to accept incoming connections via a home address and port previously registered in one of the Freedom System's wormholes.

My specific contribution to this paper was the proposed extension to the Freedom System (Sect. III.A of this paper) which enables a Freedom client to seamlessly roam among IP subnetworks and media types whilst being untraceable. By untraceable in the context of the licentiate thesis we mean the capability of a mobile node to conceal the relation between location and personal identifiable information from third parties whilst the user is on the move.

Paper #2

- **Location Privacy in IPv6: 'Tracking binding updates'.**
  Alberto Escudero-Pascual
  Tutorial at Interactive Distributed Multimedia Systems (IDMS2001). Hosted in co-operation with ACM SIGCOMM and SIGMM. Lancaster, United Kingdom. 4th September 2001.

The paper was presented as part of the MobileIPv6 tutorial held during the Interactive Distributed Multimedia Systems (IDMS2001) workshop in Lancaster. The paper outlines some of the changes that the next generation protocol has introduced and shows the location privacy threat by describing how eavesdroppers in the network will be able to identify packets that belong to a particular node and track its movements.

The paper reflects on three proposals that try to enhance privacy with respect to the level of privacy achieved: Privacy Extension for Stateless Address Configuration [18], Privacy extension to MobileIPv6 [19] and Privacy extension in Hierarchical MobileIPv6 [20].

The novel contribution of this paper is the concept of "unobservable pseudo random interface identifier" [Sect. II. C.1 of this paper]. that considers as a criteria to improve privacy when a solution makes it less obvious (observable) that a given user wants to protect their privacy.

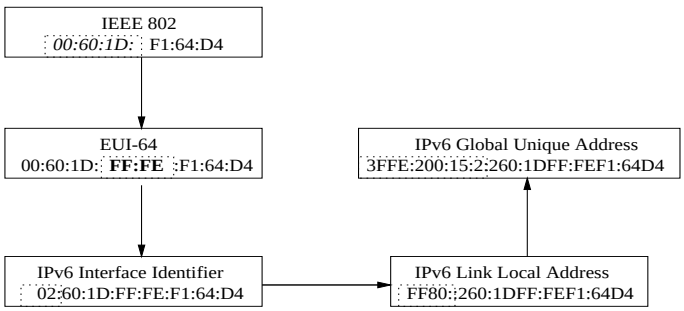

Figure 3: Generation of a global unique IP_{v6} interface identifier. The IP_{v6} address generated via Stateless Autoconfigur ation contains the same interface identifier regardless of the location the mobile node is attached to the Internet.

To the knowledge of the author, the paper is the first presentation of the limitations of RFC3041 [18] in terms of observability of privacy preferences, i.e. there are scenarios where it is possible to determine that an interface identifier has been generated as the result of an user's privacy preference.

Paper #3

- **Requirements for unobservability of privacy extension in IPv6**.
  Alberto Escudero-Pascual
  Radio Vetenskap 2002. Stockholm, Sweden. June 2002, pp. 58

After a description of the privacy concerns of the stateless address autoconfiguration mechanism for IPv6 [6] (IPv6 addresses generated via stateless autoconfiguration contain the same interface identifier (IID) regardless of the location the mobile node is attached to the Internet), the paper examines the limitations of the proposed privacy extension [18] or RFC3041 to address autoconfiguration in IPv6.

The contribution of this paper is double, the paper shows the privacy implications of the universal/local bit of the current IPv6 addressing architecture and presents a set of suggested changes to enhance privacy and secondly studies different scenarios where a third party will be able to determine if the interface identifier of a certain node has been generated as the result of using RFC3041 or not.

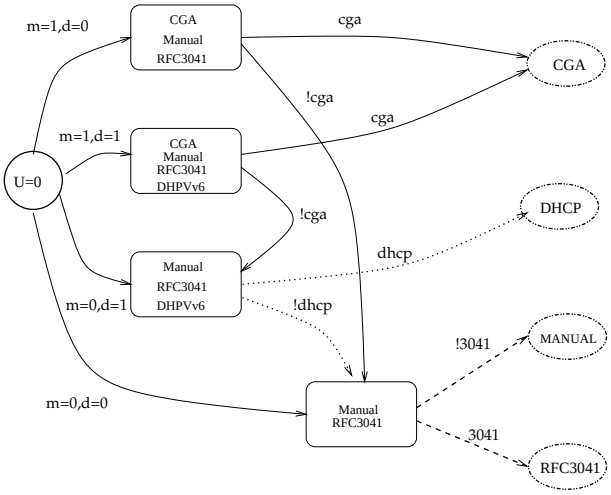

Figure 4: Possible scenarios for privacy extension observability. **m** = using mobility, **d,dhcp** = running DHCPv6, **cga** = using CGA addresses, **3041** = using privacy extension

The paper, written after a discussion in the IETF IPng mailing list [21], argues that although the *u bit* in the current Internet Identifier (IID) definition is used to indicate whether or not the IID can be considered globally unique, the *u bit* zero value can reveal under certain scenarios the fact that certain user wants to protect his or her privacy [Fig. 4].

Paper #4

- **Privacy enhanced architecture for location based services in the next generation wireless networks.**
  Alberto Escudero-Pascual
  11th IEEE Workshop on Local and Metropolitan Area Networks. LANMAN2002. Stockholm. Sweden. August 2002, pp. 169-172.

The paper proposes a privacy enhanced location based service (PE-LBS) architecture which allows a mobile node to request location based services via a proxy server hiding the network location of the mobile device while providing service accountability.

The papers outlines an architecture that makes use of XML Encryption [22] and Simple Access Object Protocol [23] so as to implement a *MIX-based SOAP Dispatcher*. The architecture enables a location based services proxy to act as a "mix" [24] by buffering and changing the sequence of the service requests, thus a mobile device can use a chain of PE-LBS proxies configured as a "mixing network" to forward location based service requests and that these functionalities can be done independently of the specific transport network.

The architecture does not employ new cryptographic techniques or protocols. However, I believe the application of these known techniques is novel and suitable for the next generation (3G) mobile phone platform.

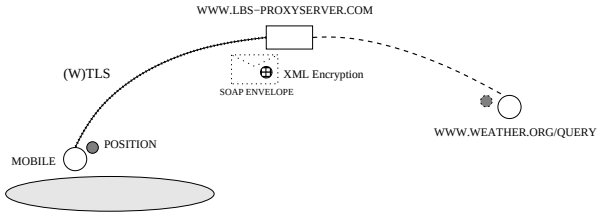

Figure 5: SOAP Request via PE-LBS proxy. The proxy acts as on behalf of user and conceals the personal identifiab le information from the location based-services provider.

The main components of the architecture is the LBS Proxy Server, responsible of processing SOAP (message envelope) requests and generate responses. When the SOAP request is received by a server, it gets bound to the class specified in the request. The proxy server works as a SOAP Dispatcher, by determining which class should handle a given request, and loading that class, if necessary. The SOAP server acts as an intermediary between a SOAP client and the requested service provider [See Fig. 5].

The privacy enhanced location based service proxy acts as a intelligent software agent that takes into consideration the privacy risks associated with the use of agents [25,26].

Paper #5

- **Role(s) of a proxy in location based services.**
  Alberto Escudero-Pascual and G.Q. Maguire Jr.
  13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications. PIMRC2002. Lisbon. Portugal. September 2002, Vol.3 1252-125, Vol.3 1252-1257

This paper is an extension of the LANMAN2002 paper, in this paper we examine a number of roles that a proxy server can play in Location Based Services and how it can be used to provide protection of personal identifiable information. In order to illustrate our approach the paper includes a description of how we have applied our privacy model to location information obtained from a Global Positioning System receiver.

The privacy enhanced location based service (PE-LBS) architecture is composed of six functional independent modules which allows a mobile node to request location based services via a proxy server: location acquisition hardware, XML data record parser, XML service request (SOAP), transport module, location-based services proxy and service modules [See Fig. 6].

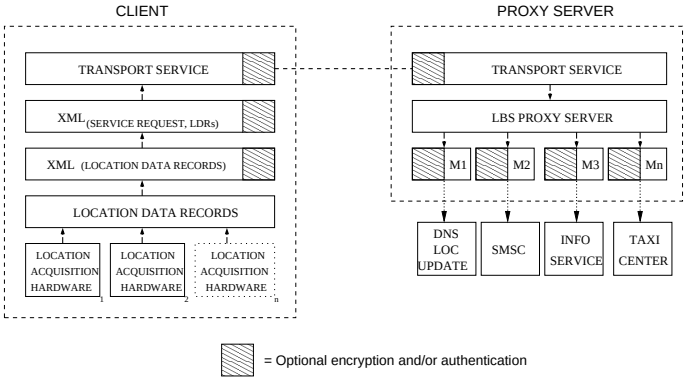

Figure 6: PE-LBS Architecture

My individual contribution of this paper is to show that by taking advantage of the extensibility and flexibility of XML we can implement the main functions of a mix network and extend the set of privacy enhanced location based services while still hiding the mobile node's network and *physical location* as desired.

Paper/Journal #6

- **The hazards of technology-neutral policy: questioning lawful access to traffic data.**
  Alberto Escudero-Pascual and Ian Hosein
  Communications of the Association for Computer Machinery (CACM) Journal. Accepted on the 5th September 2002 and reviewed the 19th October 2002.

This is the first of two papers that deal with the intersection of technology and policy development. The paper shows how the initiatives to update traditional powers of investigation involving technology do not always reflect the sensitivities raised by the current technological environment. After a review of the common policy initiative [27,28,29,30] to extend traditional powers of lawful access to communications traffic data, the paper presents some of the data that may qualify as traffic data from these communications infrastructures, to show the varying level of details that can be derived from this data.

The paper investigates two worrying trends. First, governments are updating their legislative frameworks to deal with new communications infrastructures; but they are tending towards ambiguous, or technology-neutral terminology, particularly in defining traffic data. Second, we have shown that 'traffic data' differs for each communications infrastructure and protocol, and the amount of information that can be deduced from this information increases as we look to more sophisticated communications media than the POTS. The policy language developed under POTS and sustained through 'technology-neutral' policy intentions now gives law enforcement agencies access to highly sensitive data; but only under the protections afforded to the more benign POTS procedures. In fact, 'traffic data' appears to be more 'interaction data' in which we can learn the details of an individuals intentions, thoughts, and interests; and in a sense is more sensitive than the contents of communications.

The main contribution of the paper is to show, by presenting some of the data that may qualify as traffic data, how the sensitivity of the data collected changes due to the different traffic data *granularity.* Based on our study we propose that lawful access policies must be technology-specific, and as a result governments must consider protecting the right of privacy of an individual's traffic data equally to that of communications.

My individual contribution was to investigate four different sources of traffic data and show how traffic data changes depending on the infrastructure. Some of the data presented in the paper was obtained from the *Big Brother System* [32] that was built when the Kista - IT University wireless network was being designed in October 2000. Initially designed as a networking tool to help us with the positioning of the wireless access points. *Big brother* was a monitoring system that detects the movements of the wireless users at the Kista IT-University.

Paper #7

- **Privacy in mobile Internet in the context of the European Union data protection policy.**
  Alberto Escudero-Pascual
  Internet Society's 12th Annual INET Conference. Internet Crossroads: Where Technology and Policy Intersect (INET2002). Washington DC, USA. June 2002.

The paper starts introducing how 'mobility' is supported in IPv6 and introduces the key elements of The European Commission proposal for a Directive (COM(2000)385) (now European Directive 2002/58/EC) on 'processing of personal data and protection of privacy in the electronic communication sector'.

After the technical and legal overview we discuss the difficulties of applying the definitions provided by the Directive to certain technology such as mobility in IPv6. The paper shows the kind of information items that are required to be in transit in the network to allow a mobile node to seamlessly communicate on the move and how difficult is to classify these data following the European Directive definitions of location and traffic data.

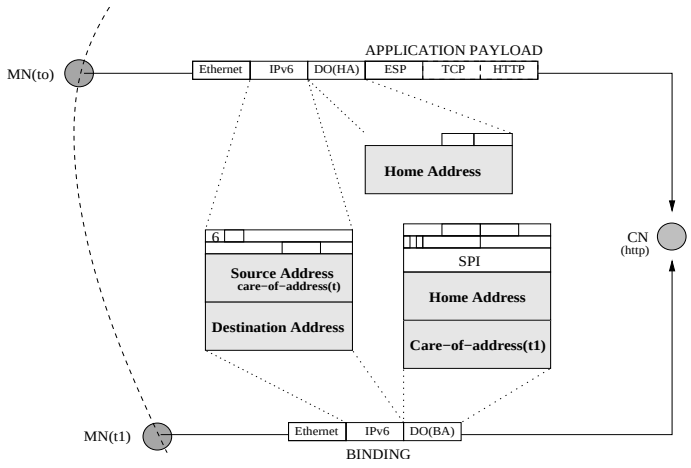

Figure 7: Mobility/Location Information embedded in IPv6 headers. Location data is embedded in "Traffic Data".

The contribution of the paper is to illustrate, by presenting a concrete technical scenario involving mobility support in IPv6 [See Fig. 7], that classifying and defining data by traditional means and ways without taking into account the Internet's multi-layered architecture might lead to an insufficient level of privacy protection for certain sensitive data.

Paper #8

- **Privacy for location data in Mobile Networks.**
  Alberto Escudero-Pascual, Thijs Holleboom, and Simone Fischer-Huebner
  Karlstad, Sweden. Nordsec2002. November 2002, pp. 220-232.

The last paper brings together many of the issues covered by my previous papers.

After a brief introduction to the three interrelated areas covered in the paper: mobility in IP networks, privacy protection for location data introduced in the new European Union data protection directive, and to means of protecting privacy by technology, we introduce the concept of *co-located displacements in MobileIP* and show how the home agent will be able to determine whether or not a set of mobile nodes move in a co-located fashion.

The paper shows that traffic data in MobileIP-based networks can also contain sensitive information about the relative position and co-location of two (or more) mobile nodes, and thus this data also needs high level of privacy protection.

Finally the paper also shows how two privacy-enhancing technologies should be applied to technically enforce legal privacy requirements of Article 9 of the European Directive 2002/58/EC for location data:

- **Mix-nets** based architectures as an effective mean for anonymising location data (requirements of Article. 9 paragraph 1).
- The **Platform for Privacy Preferences (P3P) Protocol** as a mean for enforcing the privacy principle of informed consent for location data, and also for allowing users to later revoke their consent (requirements of Article 9 paragraph 1 and 2).

3 Conclusions and future work

This dissertation has concentrated on examining new emerging privacy challenges concerning the next generation Internet. Three privacy areas have been identified and a set of technical and legal recommendations has been included in each of the related papers. The use of unique identifiers, the linkability between location and user's personal identifiable information and the legal treatment of the Internet traffic in technology-neutral regulations are the three key areas for Internet privacy examined in this dissertation.

Regarding the use of unique identifiers we have shown how eavesdroppers will be able to identify and track packets that belong to a particular node [Paper #2] and the limitations of the privacy extension for stateless address autoconfiguration which fail to provide privacy [Paper #3].

We have shown two different architectures [Papers #1, #4, and #5] that provide unlinkability between user's personal identifiable information and location information. The extensions to the Freedom System [Paper #1] enables seamless mobility while location privacy and the privacy enhanced architecture for location based services (PE-LBS) [Papers #4 and #5] describes a suitable system design that can be integrated in the next generation (3G) mobile phone platform.

The third key area is covered by a critical review of the policy initiatives to extend traditional powers of lawful access to communications traffic data [Paper #6] and the European Union Data Protection Telecommunications Directive [Paper #7].

3.1 Legal recommendations

One important goal of this research work was also to provide recommendations to regulators so that this technology and its deployment can meet the specific legal obligations of the new European Union data protection directive.

Inline with some of the results of this dissertation, on the 30th May 2002 the Article 29 Data Protection Working Party published a report titled: "Opinion 2/2002: on the use of unique identifiers in telecommunication terminal equipments: the example of IPv6". The Article 29 Data Protection Working Party is an independent advisory body on data protection and privacy set up under Article 29 of Directive 95/46/EC. The Article 29 Data Protection Working Party report [Appendix A] refers to the results of [Paper #2] of this dissertation to draw the following conclusions:

> The new IPv6 protocol allows stable connections, with maintenance of the same address, even when a terminal is moving on the network. Security and confidentiality aspects are at stake here, as there is a risk of identification of location data of this mobile node.[...] It is now widely recognized that IP address - and a fortiori a unique identification number integrated in the address can be considered as personal data in the sense of the legal framework.

Another contribution of my research to the regulatory bodies took place on the 27th November 2002 during the European Union Forum on Cybercrime. My full contribution [Appendix B] to the Forum can be summarized as follows:

> The information included in the Internet Protocol headers plus the mobile terminal locations can determine - with high precision - our human interactions, interests and behavioral habits. Therefore, 'location data' should be considered to be just as sensitive as 'content data' due to the categories of information that can be extracted from location data sets.

3.2 Future work

Future work in the three different areas that are covered in this dissertation have been identified.

- **Related to unique identifiers**

The role of a Cryptographically Generated Address (CGA) in identity management and session untraceability. A Cryptographically Generated Address was introduced to solve the problem of address ownership in MobileIP and neighbor and router discovery [33,35,34]. The Cryptographically Generated Address has a strong cryptographic binding with a public key and is obtained by means of a one-way hash function.

For example in the case of Statistically Unique and Cryptographically Verifiable (SUCV) Identifiers [36,37,38] the CGA addresses are created as follows:

$$\begin{aligned} CGA_{128} &= Prefix_{64} + CGIID_{64} \\ CGIID &= f(PK, j) \end{aligned} \tag{1}$$

where f in (1) is a function that computes the least significant 64 bits of SHA-1 hash of a Public Key (PK) concatenated with a 16 bit counter (j) and sets the universal bit (u) to zero.

This scheme enables an user to probe the "ownership" of certain Cryptographically Generated Address $CGA = f(Prefix, PK, j)$ by providing a digital signature that requires the knowledge of the correspondent's private key of the public-private key pair.

I believe that the use of CGAs can not only solve the address ownership problem, but also can be used as a privacy enhancement technology. However, further investigation is needed.

- **Related to location privacy and MIX networks**

Location information is expected to play an important role in the new services available via the third generation mobile network infrastructure. Further research is needed concerning the integration of pseudonymous services in this new infrastructure that meet the legal obligations of the new European Union data protection directive. The research should not only cover the technical requirements in both terminals and infrastructure (so as to provide pseudonymous services), but also the legal basis to enforce privacy via intermediary agents in the infrastructure.

- **Related to traffic data and anonymisation**

According to the privacy principles of data minimization and data collection avoidance, both location and traffic data should be anonymized if the effort involved is reasonable in relation to the desired effect [5]. Future work needs to be done on techniques for anonymisation of traffic data compliant with the European Directive 2002/58/EC concerning the (processing of personal) data and the protection of privacy in the electronic communications sector.

References

[1] European Commission. "Next Generation Internet priorities for action in migrating to the new Internet protocol IPv6", COM(2002) 96 final, Brussels. 21st February 2002.
http://europa.eu.int/eur-lex/en/com/cnc/2002/com2002_0096en01.pdf

[2] European Parliament. "European Telecommunication New Regulatory Framework". 2000-2002.
http://europa.eu.int/information_society/topics/telecoms/regulatory/maindocs/index_en.htm#directives

[3] European Council and Parliament. "Directive of 15th December 1997 concerning the processing of personal data and the protection of privacy in the telecommunications sector", Brussels. 15th December 1997.
http://europa.eu.int/ISPO/infosoc/telecompolicy/en/9766en.pdf

[4] European Commission. "Proposal for a new directive concerning the processing of personal data and the protection of privacy in the telecommunications sector", Brussels. 12th July 2000.
http://europa.eu.int/comm/information_society/policy/framework/pdf/com2000385_en.pdf

[5] European Parliament and Council. "Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)", Brussels. 12th July 2002.
http://www.etsi.org/public-interest/Documents/Directives/Standardization/Data_Privacy_Directive.pdf

[6] S. Thomson and T. Narten. "IPv6 Address Autoconfiguration", IETF's RFC 2462. December 1998.

[7] W. Grossman. "Conflicting Issues: Security and Privacy", The Feature. 27th August 2001.
http://www.thefeature.com/index.jsp?url=article.jsp?pageid=12550

[8] C. Macavinta. "Internet protocol proposal raises privacy concerns", CNET tech news. 14th October 1999.
http://news.com.com/2100-12-231403.html?tag=rn

[9] S. Deering and B. Hinden. "Statement on IPv6 Address Privacy". 6th November 1999. http://playground.sun.com/pub/ipng/html/specs/ipv6-address-privacy.html

[10] A. F. Westin. "Privacy and Freedom", Atheneum Press, New York, USA. 1967.

[11] M. Köhntopp, A. Pfitzmann et al. "Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology". Workshop on Design Issues in Anonymity and Unobservability, 2000
http://www.koehntopp.de/marit/pub/anon/Anon_Terminology.pdf

[12] A. Escudero. "Anonymous and untraceable communications in mobile internetworking", Department of Microelectronics and Information Technology, Royal Institute of Technology, Sweden, Licentiate Thesis, ISSN 1403-5288. May 2001.

[13] I. Goldberg. "A pseudonymous communications infrastructure for the internet". PhD disseration. Fall 2000
http://www.isaac.cs.berkeley.edu/~iang/

[14] P. Boucher, A. Haystack, and I. Goldberg. "Freedom System 2.0 Architecture", Zero Knowledge System's White Papers. 2000
http://www.freedom.net/info/whitepapers

[15] A. Fastened, D. Kesdogan and O. Kubitz. "Analysis of security and privacy in MobileIP", 4th International Conference of Communications Systems Modeling & Analysis, Nashville, USA. 1996.

[16] T. Lopatic. "Diplomarbeit Konzeption und prototypische Implementierung einer Erweiterung des Mobile Internet Protokolls", Master Thesis. November 1996
http://www.dbs.informatik.uni-muenchen.de/~lopatic/thesis.ps

[17] M. Reed, P. Syverson and D. Goldschlag. "Anonymous Connections and Onion Routing", Naval Research Laboratory Research Papers, 1998

[18] T. Narten and R. Draves. "Privacy Extensions for Stateless Address Autoconfiguration in IPv6", IETF's RFC 3041. January 2001.

[19] C. Castelluccia, "A Simple Privacy Extension for MobileIPv6", IETF's Internet Draft. February 2001.
http://www.inrialpes.fr/planete/people/ccastel/
draft-castelluccia-mobileip-privacy-00.txt

[20] H. Soliman, C. Castelluccia, K. El-Malki, and L. Bellier. "Hierarchical MIPv6 mobility management", IETF's Internet Draft. February 2001.
http://www.inrialpes.fr/planete/people/ccastel/
draft-ietf-mobileip-hmipv6-05.txt

[21] IETF's IPng Mailing list. "Next steps on Reserving bits in RFC 2473 Interface IDs - Thread discussion". 12th March 2002.
ftp://playground.sun.com/pub/ipng/mail-archive/ipng.200203

[22] W3C. "XML Encryption Syntax and Processing", Working Draft. 18th October 2001.

[23] W3C. "Simple Object Access Protocol (SOAP) 1.1". Technical Report. May 2000.

[24] D. Chaum. "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms". Communications of the ACM (24)2, pp. 84-88, 1981.

[25] European Union's Data Protection Working Group. "Common Position on Intelligent Software Agents". Norway. April 1999. http://www.datenschutz-berlin.de/doc/int/iwgdpt/agent_en.htm

[26] L. Bygrave. "Electronic Agents and Privacy: A Cyberspace Odyssey 2001", International Journal of Law and Information Technology,vol. 9, pp. 275-294. 2001.

[27] Ministers of the European Union. "Global Information Networks: Ministerial Declaration". Bonn, European Union. July 1997.

[28] United States' Delegation. "Discussion Paper for Data Preservation Workshop", Tokyo, G8 Conference on High-Tech Crime. May 2001

[29] Council of Europe. "Convention on Cybercrime Explanatory Report", adopted on November 8, 2001
http://conventions.coe.int/

[30] P. Taylor. "Issues Raised by the Application of the Pen Register Statutes to Authorize Government Collection of Information on Packet-Switched Networks", Virginia Journal of Law and Technology. Spring 2001.

[31] D. Johnson, C. Perkins, and Jari Arkko. "Mobility Support in IPv6", IETF's Internet Draft. June 2002.
http://www.ietf.org/internet-drafts/
draft-ietf-mobileip-ipv6-18.txt

[32] S. Andersson. "På KTH utvecklas teknik att stoppa övervakning". NyTeknik. November 2000.

[33] P. Nikander. "An Address Ownership Problem in IPv6", IETF's Internet Draft. February 2001.
http://www.tcm.hut.fi/~pnr/publications/
draft-nikander-ipng-address-ownership-00.txt

[34] M. Roe, T. Aura, G. O'Shea and J. Arkko. "Authentication of Mobile IPv6 Binding Updates and Acknowledgments", draft-roe-mobileip-updateauth-02 (work in progress), February 2002.

[35] J. Arkko, T. Aura, J. Kempf. V. Mantyla, P. Nikander, and M. Roe. "Securing IPv6 Neighor and Router Discovery", WiSe 2002, Atlanta, Georgia, USA. September 2002. http://www.tcm.hut.fi/~pnr/publications/WiSe2002-Arkko.pdf

[36] G. Montenegro and C. Castelluccia. "SUCV Identifiers and Addresses". IETF's Internet Draft. July 2002. http://www.inrialpes.fr/planete/people/ccastel/draft-montenegro-sucv-03.txt

[37] G. Montenegro, and C. Castelluccia, "Statistically Unique and Cryptographically Verifiable (SUCV) Identifiers and Addresses", NDSS 2002. February 2002.

[38] C. Castelluccia and G. Montenegro, "Securing Group Management in IPv6 with Cryptographically Generated Addresses", draft-irtf-gsec-sgmv6-00 (work in progress), April 2002.

Collection of Papers

1) A. Escudero, M. Hedenfalk, and P. Heselius, Location Privacy in Mobile Internet - An extension to Freedom Network. Internet Society Conference (INET2001). Stockholm, Sweden. June 2001. `http://www.isoc.org/isoc/conferences/inet/01/CD_proceedings/T06/inet2001-escuderoa-t06.pdf`
2) A. Escudero, Location Privacy in IPv6: 'Tracking binding updates'. Tutorial at Interactive Distributed Multimedia Systems (IDMS2001). Lancaster, UK. September 2001.
3) A. Escudero, Requirements for unobservability of privacy extension in IPv6. Radio Vetenskap 2002. Stockholm, Sweden. June 2002, pp.58.
4) A. Escudero, Privacy enhanced architecture for location based services in the next generation wireless networks. 11th IEEE Workshop on Local and Metropolitan Area Networks(LANMAN2002). Stockholm, Sweden. August 2002, pp. 169-172
5) A. Escudero and G.Q. Maguire Jr., Role(s) of a proxy in location based services. 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC2002). Lisbon. Portugal. September 2002, Vol3 pp. 1252-1257. ©IEEE
6) A. Escudero and I. Hosein, The hazards of technology-neutral policy: questioning lawful access to traffic data. To appear in Communications of the Association for Computer Machinery (CACM) Journal. Accepted 5th September 2002 - Reviewed 19th October 2002. ©ACM
7) A. Escudero, Privacy in mobile internet in the context of the European Union data protection policy. Internet Society Conference (INET2002). Washington DC. USA. June 2002. `http://inet2002.org/CD-ROM/lu65rw2n/papers/t07-a.pdf`
8) A. Escudero, T. Holleboom, and S. Fischer-Huebner, Privacy for location data in Mobile Networks (NORDSEC2002). Karlstad, Sweden. November 2002, pp. 220-232

N.B.
- All the papers have been reformatted and renumbered as sequential pages from their original versions to ensure consistency of page layout and numbering within this thesis.
- The papers have been reformatted using the IEEEtran LaTeXclass that provides formatting for authors of the Institute of Electrical and Electronics Engineers (IEEE) Transactions Journals.
- Paper 5) and Paper 6) are reproduced under the conditions of the copyright agreement with IEEE and ACM respectively.

PAPER #1

Alberto Escudero-Pascual, Martin Hedenfalk and Per Heselius

"Location Privacy in Mobile Internet - An extension to Freedom Network"

Internet Society Conference (INET2001)
Stockholm, Sweden
June 2001

FLYING FREEDOM: LOCATION PRIVACY IN MOBILE INTERNETWORKING

Alberto Escudero-Pascual, Martin Hedenfalk, Per Heselius
<*aep@kth.se*>, <*mhe@home.se*>, <*d97-phe@nada.kth.se*>
Royal Institute of Technology - KTH / IMIT
Electrum 204 - S 164 40 Kista
SWEDEN

Abstract - The Freedom System is a pseudonymous IP network that provides privacy protection by hiding the user's real IP addresses, email addresses, and other personal identifying information from communication partners and eavesdroppers. The following paper describes a set of protocol extensions to the Freedom System architecture to permit a *mobile node* to seamlessly roam among IP subnetworks and media types while remaining untraceable and pseudonymous.

These extensions make it possible to support transparency above the IP layer, including the maintenance of active TCP connections and UDP port bindings in the same way that $MobileIP_{v4}$ does but with the addition that the home and foreign network are unlinkable. We call this extension the *Flying Freedom System.*

I. INTRODUCTION

There are several important issues regarding security in wireless networks. As in all computer communications, these include message integrity, authentication, and confidentiality. Message integrity means that the message is transmitted without alteration, authentication means that the sending/receiving user is the one he claims to be, and confidentiality means that no one other than the intended party is able to read the transmitted message. In wireless networks, where users move between different networks and media types, another issue becomes equally important: **location privacy**. Location-aware services take advantage of the user's or terminal's location information, but what happens if the user doesn't want to be located? This means that it should be impossible to locate where a mobile user is currently working, if he/she so desires. [1]

In cellular mobile systems, such as GSM/GPRS or UMTS, it is also possible to locate users based on the cell they are in or in some cases even where within the cell the user is. In the future, a customer may choose whether this should be possible or not. Service providers could offer location privacy services as an add-on service for their customers.

A. Location privacy while seamlessly roaming

This paper presents a set of protocol extensions to the Freedom System [7] which provides similar functionality as in $MobileIP_{v4}$ [2] and also includes location privacy[4]. The Freedom System has been developed by the canadian company Zero Knowledge Systems Inc.

MobileIP allows users to move between different networks, while maintaining the same IP address. This is done by associating a care-of-address with the mobile node when it is away from home. All traffic to the mobile node is intercepted in the home network by a home agent that tunnels the data to the care-of-address.

When providing location privacy to the mobile node we need to ensure that:

- The home network should have no knowledge about which foreign network the mobile node is currently connected to.
- Similarly, the foreign or "roaming" network should have no knowledge about the mobile node's home network.
- An eavesdropper or man-in-the-middle should not be able to tell who the communicating parties are.
- In addition, all the usual communication security constraints must apply; i.e., message integrity, authentication and confidentiality.

II. A PSEUDONYMOUS IP NETWORK: FREEDOM OVERVIEW

This section is a quick overview of the Freedom System architecture and has been written with the intention of providing sufficient information to understand our protocol extensions to the Freedom

System. For a detailed look at the entities, systems and protocols that make up the Freedom System we refer the reader to the Freedom Network architecture white papers [6,7,8].

The Freedom System is a **Pseudonymous IP,** PIP, network [9]. The PIP network provides privacy protection by hiding the user's real IP addresses, email addresses and other personal identifying information from counter-parts and eavesdroppers.

The Freedom System makes it possible for a user to access the Internet without revealing any location or personal information, through the use of so called *Nyms*. The user connects to the Internet via the Freedom System that encrypts the traffic and reroutes it through special servers. Which servers to be used in the routing is determined by the user before the connection is established. Each server only knows the next and the previous proxy on the route. This way a third person eavesdropping the channel can't find out the source and destination of the connection. Since all traffic is encrypted, the content is not visible to anyone else.

The Freedom System could be seen as an overlay network composed of globally distributed servers that runs on top of the Internet. Freedom routers or **Anonymous Internet Proxies** AIP_is are the core network privacy daemons and they are in charge of passing encapsulated packets between themselves until they reach an exit node AIP_{exit} or AIP wormhole. When a certain AIP_i runs as an AIP_{exit}, it works as a traditional network address translator, NAT.

Symmetric link encryption is applied between node pairs (AIP to AIP $\{AIP_i - AIP_{i+1}\}$ and freedom-client to entry-AIP $\{FC_j - AIP_1\}$) to hide the nature and characteristics of the traffic between them.

When a freedom client with IP address IP_{FC_j} communicates with a correspondent node CN_m via a previously built virtual circuit VC_x in the Freedom System, the correspondent node sees that the traffic as coming from the wormhole IP address $IP_{AIP_{exit}}$ instead of the client's real IP address.

The client creates a virtual circuit inside the freedom network by sending a route creation packet which contains secrets S_N[1] to be shared, with each AIP_i in a chosen chain. The route create packet uses *Nested ElGamal encryption* to securely transmit the shared secrets and to ensure that each AIP_i can only read the part of the *route create packet* destined for itself. Hence, AIP_i only knows the previous AIP_{i-1} and the next AIP_{i+1} in the chain as given in the *route create packet*. The Nested ElGamal encryption is performed using the AIPs' public keys $K_{publicAIP_i}$.

The set of encryption layers (multilayer nested encryption) is called **"telescope encryption"** and it is used to provide *"freedom client-to-wormhole"* confidentiality to both route creation and data packets.

Once the route VC_x is created from the freedom client to the wormhole AIP_{exit}, the data packets travel towards the wormhole over the virtual circuit, being link decrypted, telescope unwrapped and finally link encrypted at each point.

The data is routed to the next hop by use of an **Anonymous Circuit Identifier** (ACI) mapping table. The ACIs indicate, along with a packet's implicit source address and port, the next hop in a particular route.

Data coming in over a given $[IP_{AIP_i}, Port_{AIP_i}, ACI_k]$ is first link decrypted and then telescope encrypted with the key generated from S_N. Finally the data is link encrypted and sent on its way to $IP_{AIP_{i+1}}$ with a rewritten ACI value, ACI_{k+1}.

When the ACI_k from the incoming data packet indicates that the packets from that entity need to be sent to the wormhole, the AIP_{exit} acts as a NAT_x for that connection. In this case, the wormhole will map the ACI value ($ACI_k = ACI_{exit}$) with a TCP (or UDP) local port.

A. Freedom virtual circuit example.

Let us consider a freedom client FC_j that wants to communicate with a correspondent node CN_m. The FC_j chooses a set of three AIP_i from the globally distributed Freedom AIPs $\{AIP_1 - AIP_2 - AIP_{exit}\}$. The chosen chain establishes one virtual circuit VC_x between the freedom client and the wormhole AIP_{exit}.

The freedom client negotiates a link encryption key with $AIP_1 = AIP_{entry}$ for the $\{FC_j - AIP_1\}$ link.

During the route creation process each AIP_i receives from the FC_j a unique shared secret $S_N = preKeySeed_N$ for that session. The shared secret is mapped to the ACI_k field of the incoming route create packet.

It is also during the route creation when each AIP_i is responsible for choosing a random locally unique ACI_{k+1} that will be used to send packets to the next AIP_{i+1}. The first ACI_1 in the virtual circuit chain is selected by the FC_j.

[1] The S_N, named *preKeySeed*, is a *key seed* that is used to generate keys for the three symmetric algorithms (routeCrypt, bckSymAlg and fwdSymAlg) [7].

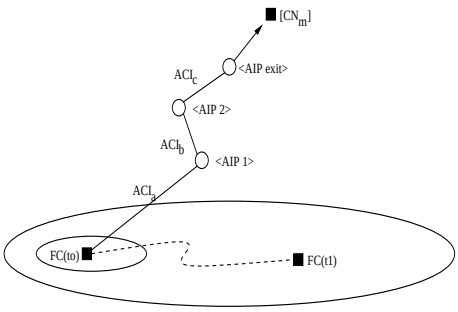

Fig. 1. Freedom Overview

In figure 1 we can see that:

- FC_j chooses ACI_a and shared secret S_A to communicate with the freedom entry AIP, AIP_1.
- After applying link decryption using the previously negotiated key with FC_j; AIP_1 knows that packets coming from IP_{FC_j} address with ACI_a can be decrypted with the key generated from shared secret S_A[2] and have to be link encrypted and forwarded to AIP_2 with rewritten ACI value, ACI_b.
- In the same way AIP_2 (after link decryption) uses S_B to decrypt packets coming from AIP_1 with ACI_b and link encrypts and forwards them to AIP_{exit} with ACI_c.
- After link decryption in AIP_{exit} the last layer of the *telescope encryption* is removed using the key derived from the shared secret S_C. AIP_{exit} also maps the packets coming from AIP_2 with ACI_c and certain port number to a local non routable IP address that will act as the source of a NAT_x session.

III. Protocol extensions to the Freedom System

Our protocol extensions to the Freedom System can be divided into two types. The first type concerns location privacy when the mobile node is run only as a client, i.e. the mobile node is only making outbound connections *(mobile client location privacy)*. The second set of extensions concerns location privacy when the mobile node also wants to act as a server accepting inbound connections from corresponding hosts *(mobile server location privacy)*.

[2]The S_A is used as key seed material to generate the key for the algorithm *(fwdSymAlg)* that is used to decrypt the corresponding encryption layer when data travels towards the AIP_{exit}.

STATE	from	to
$Before$	$FC_j[IP, Port](t_0)$ - $ACI_1(t_0)$	AIP_2 - ACI_2
$After$	$FC_j[IP, Port](t_1)$ - $ACI_1(t_1)$	AIP_2 - ACI_2

Table 1
Mappings in AIP_{entry} before and after a handover.

We have identified three different subcases for mobile client location privacy:

- **Case 0** (full route create): The mobile node sends a new **ROUTE CREATE** message after changing its point of attachment rebuilding the whole virtual circuit but keeping the same AIP_{exit} and ACI_{exit}, i.e., to preserve TCP connections and UDP port bindings.
- **Case 1** (partial route creating preserving AIP_{entry}): The mobile node sends a **ROUTE CREATE**$^{v.3}$ message[3] to the AIP_{entry} that updates the partial route. The information in the *route create packet* is used to renew the $[IP_{FC_j}(t), Port_{FC_j}(t), ACI_1(t)]$ parameters while preserving the mapping with $[IP_{AIP_2}, Port_{AIP_2}, ACI_2]$. If we represent the stages before and after a handover with t_o and t_1, then the ACI mappings in an entry AIP (AIP_{entry}) are represented as: [table 1].
- **Case 2** (partial route creating non-preserving AIP_{entry}): The mobile node sends a **ROUTE CREATE**$^{v.3}$ message upwards in the hierarchy of AIPs until the message reaches the "switching" AIP. All the routes under the switching AIP are updated preserving the higher part of the hierarchy [5].

When talking about mobile server location privacy:

- **Case** 3: The mobile server is reachable at IP_{FS_j} and $Port_{FS_j}$ allocated in a chosen AIP_{exit}, but the care-of-address of the server is not known by any AIP_i. In this case the mobile server registers an IP_{FS_j} and $Port_{FS_j}$ served by some AIP_{exit} in the freedom system. When packets arrive to that IP and port, the data travels back over the network, the information is passed along the route indicated by the ACI_k until it reaches the AIP_{entry} and link encrypted to the care-of-address of the freedom mobile server $IP(coa)_{FS_j}$. The data is transported

[3]The ROUTE CREATE and ROUTE CREATE ACK messages described in case 1-3 are not supported in current Freedom 2.x and they are part of our protocol extensions proposal.

from the AIP_{exit} to the client in the same way as data packets are transported in the normal mode of operation, i.e. the data packets are link decrypted, telescope encrypted and finally link encrypted in each AIP_i.

A. Mobile client location privacy

One of the possible scenarios looks as follows: A freedom client FC_j running an IEEE 802.11 wireless interface $IP_{FC_j}(t_0)_{WLAN}$ (while communicating with correspondent node CN_m) is moving from an indoor environment to an outdoor GPRS wireless network. A "vertical handover" is performed from one media type to another and the *mobile node* obtains a new IP address $IP_{FC_j}(t_1)_{GPRS}$ from the GPRS network.

The correspondent node is not aware of the mobility of the *mobile node* and furthermore the AIP_{exit} is also not aware of which foreign networks the mobile node is roaming in. The AIP_{exit} acts as a $MobileIP_{v4}$ home agent for the freedom client and the AIP_{entry} acts as a $MobileIP_{v4}$ foreign agent. The AIP_{entry} and the AIP_{exit} are unlinkable [3].

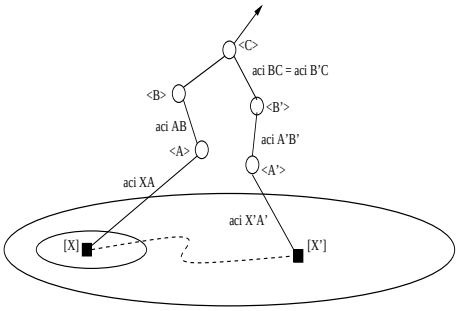

Fig. 2. Case 0: ROUTE CREATE $AIP_{exit} = AIP_{switch}$

1) Case 0: Full route create: The option for a full recreation of a route maintaining the same TCP/UDP bindings is a present feature in the Freedom 2.x architecture [Fig.2] . The *specAci* field in the **ROUTE CREATE**$^{v.2}$ packet allows the freedom client to specify the ACI_{exit} that the wormhole AIP_{exit} should use so that a route can be extended or changed using the same exit hop. This feature allows a freedom client to dynamically add a new AIP in the chain, preserving the previously allocated ACI_{exit} - NAT_x mapping.

A successful route creation is completed when the AIP_{exit} checks and validates the **Nym signature**. All packets received by AIP_{exit} from AIP_{exit-1} with a certain source port number and an ACI value (ACI_{exit}) are mapped to a local non routable IP address that will act as the source of a NAT_x session.

This reserved ACI_{exit} in the AIP_{exit} is used to identify the socket used to communicate through TCP/UDP with the corresponding host CN_m. The allocated ACI_{exit} is sent back in the **ROUTE CREATE ACK**$^{v.2}$ answer in response to the client's initial **ROUTE CREATE**$^{v.2}$ message. A **ROUTE CREATE ACK**$^{v.2}$ packet is just a data packet with the data-packet type field set to a special value. The payload carries the 2 byte ACI_{exit} number allocated in the AIP_{exit} for that session.

When the client wants to change its point of attachment, it sends a new **ROUTE CREATE**$^{v.2}$ message using the *specAci* field that is set to the $ACI_{exit}(t_0)$. The $ACI_{exit}(t_0)$ is acquired from the **ROUTE CREATE ACK**$^{v.2}$ message from the previous route creation. This way the TCP connection and UDP port bindings between the AIP_{exit} and the CN_m are preserved, and thus the connections between the mobile client and the corresponding host.

The whole route is rebuilt between the AIP_{entry} and AIP_{exit}, even where those routes are unchanged. The **Nym signature** is also rechecked at the AIP_{exit}.

From the point of view of all applications running on the freedom client the connection looks unchanged though the client IP address has changed ($IP_{FC_j}(t_0) \neq IP_{FC_j}(t_1)$).

2) Case 1: Route creating a preserving AIP_{entry}: This is the first of the proposed extensions of the Freedom System, to be able to change the point of attachment, while preserving TCP/UDP connections, but without rebuilding the whole route [Fig 3]. The freedom client gets a new IP address $IP_{FC_j}(t_1)$ (perhaps due to a move to another network) but uses the same $AIP_{entry}(t_0) = AIP_{entry}(t_1)$.

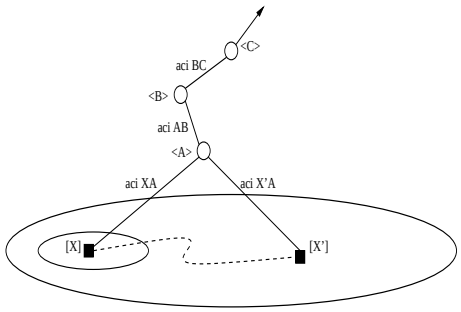

Fig. 3. Case 1: ROUTE CREATE $AIP_{entry} = AIP_{switch}$

As shown in *case 0* the current solution requires that the whole route is rebuilt (except for the socket binding in the AIP_{exit}) and that the **Nym signature** is rechecked.

Case 1 presents an alternative when the mobile node wants to keep using the same entry AIP ($AIP_{entry}(t_0) = AIP_{entry}(t_1)$). In this case, the whole route does not need to be rebuilt.

In order to update the route binding for the mobile node, the AIP_{entry} needs to be notified that:

- the freedom client has a new IP address $IP_{FC_j}(t_1)$.
- the freedom client already has had a route binding for the old IP address $[IP_{FC_j}(t_0), Port_{FC_j}(t_0), ACI_1(t_0)]$.

The mobile node first has to exchange a new shared secret with the entry AIP ($AIP_{entry}(t_1)$) to be able to establish new link encryption between the freedom client and the entry AIP. $\{FC_j(t_1) - AIP_{entry}(t_1)\}$.

The mobile node then sends a **ROUTE CREATE**$^{v.3}$ message, as described in [10], that contains the old IP address $IP_{FC_j}(t_0)$, old port $Port_{FC_j}(t_0)$, old $PreKeySeed_{AIP_{entry}}(t_0)$ and old ACI $ACI_1(t_0)$. The AIP_{entry} then checks the authenticity of the message by checking that the $PreKeySeed_{AIP_{entry}}(t_0)$ sent with the update is the same as the one that was previously exchanged between the client and entry AIP. ($IP_{FC_j}(t_0)$ with $ACI_1(t_0)$). If the message is verified to be correct, it then updates its route binding (uniquely identified with the $[IP_{FC_j}(t_0), Port_{FC_j}(t_0), ACI_1(t_0)]$) with the new $[IP_{FC_j}(t_1), Port_{FC_j}(t_1), ACI_1(t_1)]$ which is extracted from the **ROUTE CREATE**$^{v.3}$ header, see [table 1]

3) Case 2: Route creating a non-preserving AIP_{entry}: To generalize *case 1*, we introduce the concept of a "switching AIP", AIP_{switch} [Fig. 4].

When the mobile node changes its point of attachment (IP address), it may not want to use the same AIP_{entry}. For example, it may be impossible to use the same AIP_{entry} because it resides in a private network.

However, some $AIPs$ in the route can be the same, so the minimum route that needs to be rebuilt, is the partial route upwards to the first common AIP (AIP_{switch}).

If the mobile node selects $AIP_{switch} = AIP_{entry}$, this would behave as in *case 1*. If $AIP_{switch} = AIP_{exit}$, this case would behave as in *case 0* [table 2].

In any case the mobile node first has to perform a new key exchange with the new $AIP_{entry}(t_1)$ to be able to establish link encryption between the client with the new IP address and the entry AIP $\{FC_j(t_1) - AIP_{entry}(t_1)\}$.

Case	AIP_{switch}
Case 0	AIP_{exit}
Case 1	AIP_{entry}
Case 2	AIP_i

Table 2
AIP_{switch} depending on the case.

The client sends a **ROUTE CREATE**$^{v.3}$ message along the new specified path, up to the switching AIP, AIP_{switch}. The AIP_{switch} discovers that this is actually an update of an existing route, updates its bindings and disables the old route by sending a teardown message down the old path. If this teardown message is lost, the old route will eventually time out, since no new data will go that way.

In the same way as in *case 1* the **ROUTE CREATE**$^{v.3}$ message verifies to the AIP_{switch} that this message is authorized to update the binding represented by $IP_{AIP_{switch-1}}(t), Port_{AIP_{switch-1}}(t), ACI_{switch}(t)]$ from the values at t_0 to the new ones at t_1.

To succeed with this, the **ROUTE CREATE** $^{v.3}$ contains the old IP address and port of the next lower entity $(IP_{AIP_{switch-1}}(t_0))$- $(Port_{AIP_{switch-1}}(t_0))$, the old $PreKeySeed_{AIP_{switch}}(t_0)$ and the old ACI $ACI_{switch}(t_0)$.

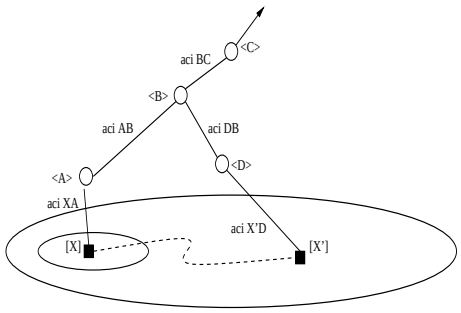

Fig. 4. Case 2: ROUTE CREATE $AIP_i = AIP_{switch}$

This means that the client must know all ACI_i used along the route. This can be accomplished by modifying the **ROUTE CREATE ACK**$^{v.3}$ message, sent back from the initial **ROUTE CREATE**$^{v.3}$ message, so that each AIP_i in the chain adds its own ACI_k number to the message before it is passed on along the route. The client already knows the IP addresses of all AIPs (AIP_i), since it is the client's responsibility to choose the chain of AIPs (AIP_i) in the first place.

If the $PreKeySeed_{AIP_{switch}}(t_0)$ is verified to be

correct, the AIP_{switch} sends a teardown message down the old route, and updates its bindings to reflect the change.

The **ROUTE CREATE**$^{v.3}$ message is similar to the standard **ROUTE CREATE**$^{v.2}$ message, in the way that new shared secrets $S_N(t_1)$ are exchanged between the client and the new set of AIPs ($AIP_i(t_1)$, $i < switch$), within each layer of the *telescope encryption*.

The multilayer encryption ensures that each secret is only known by the respective AIP_i. The client reuses the secret established with the AIP_{switch} in the initial **ROUTE CREATE**$^{v.3}$ message $S_{switch}(t_0) = S_{switch}(t_1)$.

The AIP_{switch} has to send an acknowledgment that the route actually has been updated. This message is identical to the **ROUTE CREATE ACK**$^{v.3}$ except it only contains the newly chosen ACIs (ACI_i, $i < switch$) in the partial route.

4) Switching policies: How does the client decide what AIP should be the switching one? Three possible policies are:

- Preserve as much of the old route as possible. This yields a shorter path for the **ROUTE CREATE** message, which in turn yields faster handover.
- Optimize the route length. This yields fewer hops in the route from the client to the destination.
- Change more of the route than actually needed, to increase the privacy level.

5) Entry AIP discovery: In our scenario we used a mobile client with both a wireless LAN like 802.11b and a GPRS interface. The mobile client wants to roam between different IP networks hiding the mobility from both the correspondent node and the wormhole.

The mobile node needs to know which entry AIPs (AIP_{entry}) are available in the different IP networks it is roaming in.

The mobile client determines which AIP_{entry} to use based on the following discovery procedure:

All AIPs (AIP_i) sends out an "AIP advertisement" periodically. The "AIP advertisement" message is a standard ICMP router advertisement message with a *Freedom AIP advertisement extension*, see [10]. The TTL field should be set to 1, and the destination should be 255.255.255.255 (limited broadcast).

The client can also force an advertisement by sending out "AIP solicitation" messages. The "AIP solicitation" message is a standard ICMP router solicitation message with TTL set to 1.

Which AIPs to use in the rest of the route created is determined by the user, based on the information retrieved from the freedom core servers.

One interesting feature of Freedom System that can be used to speed up handovers is that the client can create secure links between itself and *more than one* AIP_i. The $\{FC_j(t_1) - AIP_{entry}(t_1)\}$ encryption link can be established prior a new route creation is requested.

6) Handover: The mobile node performs handover when a change in point of attachment has been detected. The change can be detected either because the old connection is lost or by the client receiving agent advertisement messages from a new network. If a connection is lost then the client sends an agent solicitation message to trigger an agent advertisement message.

B. Mobile server location privacy

With this second type of protocol extension we want to allow an external node to start a connection to a *mobile server*, using the Freedom System, via an IP address IP_{FS_j} and port $Port_{FS_j}$ previously registered in the AIP_{exit}.

The AIP_{exit} acts as a home agent for the mobile server, accepting incoming connections and making the data travel back over the network to the care-of- address that the mobile server is using while moving.

The information is passed along the route indicated by the $ACIs$ until it reaches the AIP_{entry} and then it is link encrypted to the mobile server IP address. The data is transported from the AIP_{exit} to the client in the same way as data packet is transported in the normal mode of operation, i.e. the data packet is link decrypted, telescope encrypted and finally link encrypted in each AIP. The AIP_{entry} acts a foreign agent for the mobile server. The IP address of the server is in fact its care-of-address $IP(coa)_{FS_j}$.

The mobile server that wants to be reachable via the Freedom System opens a "control connection" to the AIP_{exit} and registers an IP address and port where the AIP_{exit} should listen to incoming connections. This registration is mapped with an ACI_{exit}. This ACI_{exit} binding is created by sending a **ROUTE CREATE**$^{v.3}$ message that includes the number of IP addresses and ports to be registered with the exit AIP and how those IP addresses and ports should be mapped to the remote local ports that the service is listening to on the mobile server.

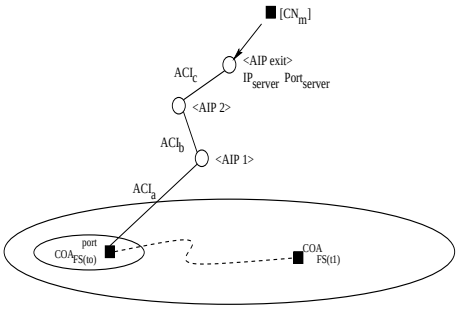

Fig. 5. Mobile Server

In the previous cases the AIP_{exit} maps the mobile node port number and ACI_{exit} with a random port selected by the NAT for the outbound connection.[4] In our new case the **ROUTE CREATE**$^{v.3}$ message would explicitly inform the AIP_{exit} that inbound connections to the mobile server's "home address" with *port X* should be mapped to ACI_{exit} and *port X'*.

Each AIP_i keeps a *control timer* for each virtual circuit that is reset when data is transmitted. When the timer expires a *teardown message* is generated to destroy the virtual circuit. If a corresponding host tries to connect when the route has been destroyed, this would of course lead to the inbound connection being lost.

To avoid the destruction of the route we could regularly send "keep alive" traffic. The best way would be to send a **ROUTE KEEPALIVE** message regularly at some time interval suitable compared to the route teardown timer. This message is basically a normal *data packet* with a special type [10]. Data packets of this type are passed along the route to the ACI_{exit} where they are discarded. Each AIP resets their expiration timers before passing this message to the next AIP.

IV. Conclusions

Our protocol extensions of the Freedom System permits a mobile client to seamlessly roam among IP subnetworks and media types while remaining untraceable. These extensions makes it possible to support transparency above the IP layer, including the maintenance of active TCP connections in the same way that $MobileIP_{v4}$ does but with the addition that the **home network and foreign network are unlinkable** [4].

The new proposed routing messages provide the mobile node the **flexibility of rebuilding partial routes** hiding the mobility associated with certain pseudo-*Nym.*

We have also introduced the possibility of having an **unlocated mobile server** roaming behind the Freedom System. The mobile server is able to accept incoming connections via a home address/port previously registered in one of the Freedom System's wormholes.

[4] In order to guarantee that IP and port are unique for the NAT, the AIP_{exit} allocates one fake non routable IP address per ACI_{exit}. The mobile node local port number and ACI_{exit} are mapped to IP_{fake} and a random port number. IP_{fake} and $Port_{IP_{fake}}$ are used as source of *requests connections* to the NAT.

About the Authors

Alberto Escudero <alberto@it.kth.se>, is graduate student at the Institution of Microelectronics and Information Technology. (formerly Teleinformatics). KTH. Sweden. As a researcher in IMIT, Alberto is currently focusing on *location privacy* in mobile internetworking and privacy protection of personal information and over the last year was involved in the IT University - Kista Wireless Network, FlyingLinux.NET and The Big Brother "storebror" Project. http://www.it.kth.se/~aep

Martin Hedenfalk <mhe@home.se>, is an undergraduate student in Electrical Engineering at the Royal Institute of Technology, Stockholm, Sweden. Martin has many years of experience with GNU/Linux and has developed networking software for the Open Source community and currently combines his studies by working part time as a Linux consultant.

Per Heselius <d97-phe@nada.kth.se>, is an undergraduate student in Computer Science at the Royal Institute of Technology in Stockholm, Sweden. Per is interested in developing systems for Computer Security, Cryptography and Computer Communications and is currently finishing the last year of his M.Sc. in Computer Science and Engineering.

References

[1] Forrester Report. *"Surviving the privacy revolution"*. February 2001

[2] Perkins, C. *"IP Mobility support, RFC 2002"*. 1996

[3] Pfitzmann, A., Koehntopp, M. *"Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology"*. 2000 http://www.koehntopp.de/marit/publikationen/

[4] Fasbender, A., Kesdogan, D., Kubitz, O. *"Variable and Scalable Security: Protection of Location Information in Mobile IP"*.

[5] Forsberg, D., Malinen, J.T., Maline, J.K., Weckstrm, T. *"Dynamics - HUT Mobile IP a Technical Document"*. 1999
`http://www.cs.hut.fi/Research/Dynamics/documents.html`

[6] Boucher, P., Shostack, A., Goldberg, I. *"Freedom System 2.0 Architecture"*. 2000
`http://www.freedom.net/info/whitepapers`

[7] Goldberg, I., Shostack, A. *"Freedom Network 1.0 Architecture and Protocols"*. 1999
`http://www.freedom.net/info/whitepapers`

[8] Back, A., Goldberg, I., Shostack, A. *"Freedom 2.0 Security Issues and Analysis"*. 2000
`http://www.freedom.net/info/whitepapers`

[9] Goldberg, I. *"A pseudonymous communications infrastructure for the internet"*. Fall 2000
`http://www.isaac.cs.berkeley.edu/~iang/`

[10] Hedenfalk, M., Heselius P., Escudero A. *"Location privacy protocol extensions to the Freedom System"*. April 2001
`http://www.it.kth.se/~aep/publications`

PAPER #2

Alberto Escudero-Pascual

"Location Privacy in IPv6: 'Tracking binding updates'"

Tutorial at Interactive Distributed Multimedia Systems (IDMS2001)
Lancaster, United-Kingdom
September 2001

- LOCATION PRIVACY IN IP_{V6} - "TRACKING THE BINDING UPDATES"

Alberto Escudero-Pascual, <aep@kth.se>

Abstract - The following paper shows some of the changes that IP_{v6} has introduced in order to meet the needs of the mobile Internet, and how it will be possible for eavesdroppers in the network to identify packets that belong to a particular node and track its movements.

Three proposals that try to prevent this kind of location information leakage are presented within the level of privacy achieved. The first one deals with stateless address autoconfiguration and the generation of random interface identifiers; the second tries to hide the home address of the mobile node from third parties using a temporary mobile identifier and the last one uses hierarchical $MobileIP_{v6}$ basic mode to hide the link care of address to correspondent nodes by using the regional care-of-address.

The main contribution of the author in this paper is the concept of *unobservable pseudo random interface identifier* which enhances the privacy extension for stateless address autoconfiguration.

I. Introduction

There are several important issues regarding security in any kind of communications. These include message integrity, authentication, and confidentiality. Integrity means that the message is transmitted without alteration, authentication means that the sending/receiving user is the one they claim to be, and confidentiality means that no other one than the intended party, is able to read the transmitted message.

The IETF IP Security Protocol Working Group is responsible for the development of the mechanisms to protect client protocols of IP. The $IPSEC\ WG$ has developed a security protocol in the network layer that provides cryptographic security services that supports combinations of authentication, integrity, access control, and confidentiality. The IP Encapsulating Security Payload (ESP) and the IP Authentication Header (AH) are part of the IP Security architecture described in RFC 2401 [1].

Both ESP and AH focus on the message itself (IP datagram) and make sure that a third person eavesdropping the channel can not read and/or modify the message. The IP Authentication Header seeks to provide security by adding authentication information to each IP datagram whilst confidentiality requires the use of ESP. Neither AH nor ESP hide the source and destination IP addresses of the communicating parties and hence their location.

In wireless networks, where users move between different networks and media types, another issue becomes equally important: location privacy. Location awareness services can take advantage of the possibility of gathering location information, but what happens if the user wants to conceal its location from third parties?

The paper describes some of the changes that IP_{v6} has introduced in order to meet the needs of the future Internet and the associated privacy implications.

II. Stateless Address Autoconfiguration in IPv6

Stateless address autoconfiguration defines the mechanism for a IP_{v6} node to generate an address without the need of an external DHCP server based on the interface identifier [5],[2]. The stateless approach is used when a site is not particularly concerned with the exact addresses hosts use.

To insure that all configured addresses are likely to be unique on given link, nodes run a "duplicate address detection" algorithm on addresses before assigning them to an interface. The Duplicate Address Detection (DAD) algorithm is performed on all addresses, independent of whether they are obtained via stateless or stateful autoconfiguration.

A. Interface Identifier

In the case of an Ethernet the Interface Identifier is based on the EUI-64 identifier derived from the interface's built- in 48-bit IEEE 802 address (MAC address).

The EUI-64 is formed as follows: The first three octets that corresponds to the Organizationally Unique Identifier (OUI) of the Ethernet address becomes the *company_id* of the EUI-64. The fourth and fifth octets of the EUI-64 are set to the fixed value FFFE

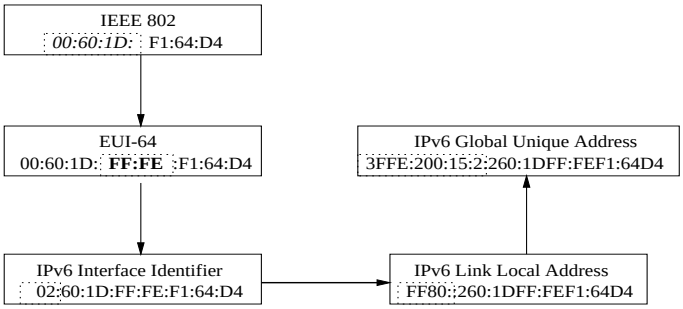

Fig. 1. Generation of a global unique IP_{v6} interface identifier

hexadecimal. The last three octets of the Ethernet address become the last three octets of the EUI-64.

The Interface Identifier is then formed from the EUI-64 by complementing the "Universal/Local" (U/L) bit, which is the next-to- lowest order bit of the first octet of the EUI-64. Complementing this bit will generally change a 0 value to a 1, since an interface's built-in address is expected to be from a universally administered address space and hence have a **globally unique value**.

A universally administered IEEE 802 address or an EUI-64 is signified by a 0 in the U/L bit position, while a globally unique IPv6 Interface Identifier is signified by a 1 in the corresponding position.

B. Global IP$_{v6}$ address and location hiding.

The link-local address of an IP_{v6} node is the result of combining the global unique interface identifier with the reserved link-local prefix FE80. The site-local and global-scope addresses are created by combining prefixes advertised in Router Advertisements and Neighbor Discovery.

The IP_{v6} address generated via Stateless Autoconfiguration contains the same interface identifier regardless of the location the mobile node is attached to the Internet.

Even when higher communication layers encrypt their payloads (for example with ESP), there is not an easy mechanism to hide the addresses in packet headers and they appear in clear, this fact makes very easy for an eavesdropper to track mobile nodes by analyzing the prefixes related with a certain interface identifier.

C. Privacy Extension for Stateless Address Configuration

Narten and Draves [3] developed a privacy extension for Stateless Address Configuration based on the idea of generating random interface identifiers periodically. They describe two approaches for the maintenance of the randomized interface identifier when stable storage is present (history scheme) or not. The use of the history scheme tries to avoid the scenario where two nodes generates the same randomized interface identifier, both detect it via DAD, but then proceed to generate identical randomized interface identifier via the same flawed random number generation algorithm.

In a real network environments the $IEEE$ 802 addresses are not random in the pure sense, some OUIs are more present than others. A better privacy protection can be achieved if the random interface identifier can not be distinguished from a *common* one. i.e. an eavesdropper can not determine if certain node is using or not the stateless address configuration privacy extension.

1) Unobservable pseudo random interface identifier: One possible solution to generate a so called "random interface identifier" that can not be distinguised from a common one is to impersonate someone else EUI-64 when not being present but a much better approach is to generate a pseudo-random interface identifier based on the statistical probability of finding certain OUIs in the working media.

The main idea is that the mobile node should keep statistical records of the presence of the different OUIs in the media and generate a random identifier based on that information. Alternatively, the router can transmit that information at regular intervals on the *all hosts* multicast address as part of the *Router advertisement* ICMP messages.

By generating unobservable random interface identifers an eavesdroppper or communicating party will not be able to determine if the IP_{v6} node is running a privacy extension or not.

A study conducted by the author at the HAL2001

OUI	%	Organization
00:20:2D	60	Lucent Technologies
00:60:1D	21	Lucent Technologies
00:30:65	8	Apple Computer
00:40:05	1	Linksys Group
xx:xx:xx	10	Other vendors

Table 1
OUIs at HAL2001

(Hackers at Large Conference) wireless network in August 2001 shows up that more othan 80% of the 580 $IEEE-802.11b$ mobile nodes run 00:02:2D (60%) or 00:60:1D (21%) Lucent Technologies interfaces [6].

III. MOBILEIPV6

A. MobileIPv6 overview

$MobileIP$ allows users to move between different networks, while maintaining the same IP address. This is done by associating a care-of-address with the mobile node when it is away from home. All traffic to the mobile node is intercepted in the home network by a home agent that tunnels the data to the care-of-address.

Packets may be routed to the mobile node using this address regardless of the mobile node's current point of attachment to the Internet, and the mobile node may continue to communicate with other nodes after moving to a new link. With $MobileIP$ the movement of a mobile node away from its home link is thus transparent to transport and higher-layer protocols and applications.

$MobileIP_{v6}$ [7] shares many features with $MobileIP_{v4}$, but the protocol is now fully integrated into IP_{v6}. $MobileIP_{v6}$ works in a similar way as $MobileIP_{v4}$ does when using mobile node co-located care of address mode and route optimization.

As in $MobileIP_{v4}$ the mobile mode is responsible for discovering its current location. When the mobile mode is attached to its home link it works as a fixed host and when roaming in a foreign network, it must acquire a co-located care of address and notify this address to the home agent.

$MobileIP_{v6}$ on the other hand also includes mechanisms that allows the mobile node to inform to a selected IP_{v6} correspondent hosts of its care-of-address so packets from the correspondent hosts can be redirected straight to the mobile node instead of using the home agent as an intermediary.

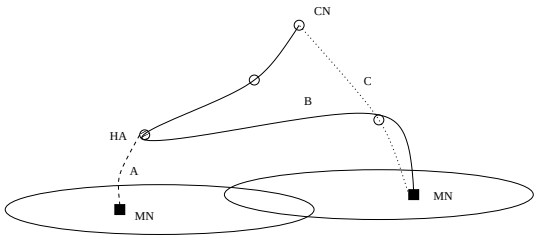

Fig. 2. MobileIPv6 triangle routing

The association between a mobile node's home address and care-of address is known as a **"binding"** for the mobile node. A mobile node typically acquires its care-of address through stateless or stateful (e.g., $DHCP_{v6}$) Address Autoconfiguration.

While away from home, a mobile node registers its care-of address with a router on its home link, requesting this router to function as his *home agent*. This binding registration is done by the mobile node sending to the home agent a packet containing a *Binding Update* destination option; the home agent then replies to the mobile node by returning a packet containing a *Binding Acknowledgement* destination option.

The Binding Update and Binding Acknowledgement destination options, together with a *Binding Request* destination option, are also used to allow IP_{v6} nodes communicating with a mobile node, to dynamically learn and cache the mobile node's binding.

B. Learning and caching binding updates

Before sending a packet to any IP_{v6} destination, a node checks its *cached bindings* for an entry for the packet's destination address. If a cached binding for this destination address is found, the node uses an IP_{v6} *Routing header* to route the packet to the mobile node by way of the care-of address indicated in the cached binding. No IP_{v6} encapsulation is required, at destination the mobile node sees the home address in the routing header and gets the packet.

If the sending node has no cached binding for the destination address, the node sends the packet normally to the home address without any *Routing header* and the packet is subsequently intercepted and tunneled by the mobile node's home agent.

The correspondent nodes can also send a *Binding reques*t to the mobile node home address in order to obtain its care-of-address.

C. Privacy extension to MobileIPv6

In $MobileIP_{v6}$ the home address of a mobile node is included in cleartext in the packets it sends and receives. Also the packets sent by a correspondent node to a given mobile node contains a routing header that includes the mobile home address. Any eavesdropper within the network can easily identify packets that belong to a particular mobile node and track mobile movements.

Castelluccia [4] proposes to assign each mobile node a TMI (Temporal Mobile Identifier). This TMI, a random 128-bits sequence, is used by the mobile node's home agent and correspondent nodes to identify the mobile node.

When the mobile node initiates the communication, packets sent and received by a mobile node will contain its TMI instead of its home address. As a result, the mobile identity (home network) is hidden from the correspondent node and from potential eavesdroppers in the network.

The second scenario is that the correspondent node initiates the communication, in this case the correspondent node knows the mobile home address (identity) by definition. If a mobile node wants to hide its mobility, i.e. its care-of address, from a particular correspondent node, it must not send any binding update to this correspondent node and use bi-directional tunneling. As a result the packets that are sent to the mobile node are addressed to its home address and encapsulated by the home agent to its current care-of address.

When the eavesdropper is located somewhere along the route between the home agent and the mobile node it is possible to identify and track the mobile movement by looking at the inner packet. Therefore the packets that are sent between the mobile node and its home agent should be encrypted.

In the case that the correspondent node initiates the communication and the mobile node decides to use route optimization, the mobile node sends a binding update to its correspondent node that contains the TMI in the home address option, and the actual home address is encoded in a newly defined binding update suboption.

To preserve privacy the binding update must be encrypted and the security association should be indexed with the TMI, not the home address. The correspondent node uses the binding update to bind the TMI with the Home Address and the care-of address.

Subsequent packets between the mobile node and the correspondent node will contain the TMI in the home address option and in the routing header extension instead of the actual home address.

Castellucia's $MobileIP_{v6}$ privacy extension aims to make it more difficult for an eavesdropper to identify the packets belonging to a particular mobile node (home address) when roaming in a foreign network. The mobile node home address and care-of-addresses remain unlinkable by hiding the mobile node home address from third parties.

Unfortunately this privacy extension is not enough against an attacker that is able to perform simple traffic analysis which will try to correlate binding updates carrying a TMI and the binding carrying the real home address.

IV. Hierarchical MobileIPv6

H. Soliman et al.[8] introduce some extensions for $MobileIP_{v6}$ and neighbor discovery for supporting a hierarchical mobility management model in IP_{v6}, utilizing a new node called the Mobility Anchor Point.

In $MobileIP_{v6}$ there are no Foreign Agents, but there is still the need to provide a central point to assist with handoffs. Similar to $MobileIP_{v4}$, $MobileIP_{v6}$ can benefit from reduced mobility signalling with external networks by employing a local hierarchical structure. For this reason a new node, called the Mobility Anchor Point (MAP), is used and can be located at any level in a hierarchical $MobileIP_{v6}$ network including the Access Router (AR). Unlike FAs in IP_{v4}, a MAP is not required on each subnet.

In $HMIP_{v6}$ basic mode the MN would have two addresses, a regional care-of-address $(RCoA)$ on the MAP's subnet and an on-link care-of-address $(LCoA)$. This $(RCoA)$ is formed in a stateless manner by combining the MAP's subnet prefix received in the MAP option with the MNs interface identifier.

The $HMIP_{v6}$ basic mode is very simple in the sense that it only requires special treatment at the Mobile Nodes. The HA is unchanged. The MAP is merely a "local" HA that maps the MN's $RCoA$ to $LCoA$.

A. Privacy extension using $HMIPv6$

With HMIPv6 "basic mode", a mobile node may choose to hide its $link$ care of address $LCoA$ from its

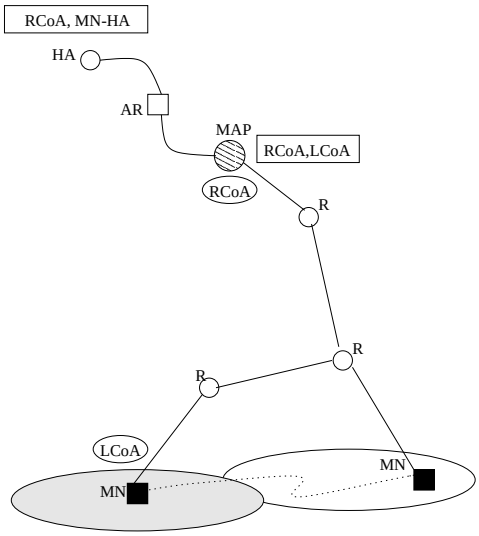

Fig. 3. $HMIP_{v6}$ basic mode

corresponding nodes and its home agent by using its $RCoA$ in the source field of the packets that it sends. As a result, the location tracking of a mobile node by its corresponding nodes or its home agent is difficult since they only know its $RCoA$ and not its actual $LCoA$.

While this feature of $HMIP_{v6}$ is clearly a marked improvement over $MobileIP_{v6}$ an eavesdropper can still determine the mobile node home address by snooping its packets.

V. Conclusions

The three proposals included in this paper provides certain levels of location privacy from correspondent nodes but not from eavesdroppers.

The privacy extension for stateless address configuration can be enhanced by generating unobservable random interface identifiers.

Future work needs to be done to conceal location information from eavesdroppers in IP_{v6} networks. Mixing techniques already used in IP_{v4} location privacy [10] can take advantage of IP_{v6} architecture.

References

[1] **S. Kent, R. Atkinson**, *"Security Architecture for the Internet Protocol"*. RFC 2041.

[2] **R. Hinden, M. O'Dell, S. Deering**, *"An IPv6 Aggregatable Global Unicast Address Format"*, RFC 2374, July 1998.

[3] **T. Narten and R. Draves**, *"Privacy Extensions for Stateless Address Autoconfiguration in IPv6"*. RFC3041, January 2001.

[4] **C. Castelluccia,** *"A Simple Privacy Extension for Mobile IPv6"*. draft-castellucia-mobileip-privacy. February 2001.

[5] **S. Thomson and T. Narten,** *"IPv6 Address Autoconfiguration"*, RFC 2462, December 1998.

[6] **A. Escudero. "***Location privacy in IP_{v6} internetworking - Pseudorandom interface identifiers"*. Hackers at large 2001. HAL2001. Twente. NL. August 2001.

[7] **D. Johnson and C. Perkins,** *"Mobility Support in IPv6"*, draft-ietf-mobileip-ipv6 v14, July 2000.

[8] **H. Soliman, C. Castelluccia, K. El-Malki and L. Bellier,** *"Hierarchical MIPv6 mobility management"*, draft-ietf-mobileip-hmipv6 v3. February 2001.

[9] **C. Perkins,** *"IP Mobility Support"*, RFC 2002, October 1996.

[10] **A. Escudero**, *"Anonymous and Untraceable Communications: Location Privacy in Mobile Internetworking"*. Licentiate Thesis. July 2001.[1]

[1]Special thanks to John Wells <wells@vt.edu> from Virginia Tech. University to make a revision of this paper.

PAPER #3

Albero Escudero-Pascual

"Requirements for unobservability of privacy extension in IPv6"

Radio Vetenskap 2002 (RVK02), pp. 58
Stockholm, Sweden
June 2002

PRIVACY EXTENSIONS FOR STATELESS ADDRESS AUTOCONFIGURATION IN IPV6 - ”REQUIREMENTS FOR UNOBSERVABILITY”

Alberto Escudero Pascual <aep@kth.se>
Royal Institute of Technology (KTH)
Kista. Sweden

Abstract - Stateless address autoconfiguration defines the mechanism for a IPv6 node to generate an address without the need of an external DHCP server based on the interface identifier. In the case of Ethernet the Interface Identifier is based on the EUI-64 identifier derived from the interface's built-in 48-bit IEEE 802 address (MAC address). The IPv6 address generated via Stateless Autoconfiguration contains the same interface identifier regardless of the location the mobile node is attached to the Internet. RFC3041 presents a privacy extension to Stateless Autoconfiguration based on the idea of generating random interface identifiers periodically.

The paper introduces the concept of “unobservability” of the privacy extension and studies in which scenarios a third party will be able to determine with high probability if a node is running RFC3041 or not. The paper shows the privacy implications of the universal/local bit of the current IPv6 addressing architecture and presents a set of suggested changes to enhance privacy.

I. BACKGROUND

This paper is divided as follows: Section 1 contains a very brief overview of the Internet Protocol Version 6, Section 2 describes the different mechanisms to generate a global scope address including the privacy extension for stateless address configuration RFC3041 [6]. Section 3 defines unobservability of the privacy extension. Section 4 introduces the level of privacy extension observability for different scenarios and finally in Section 5 we present a set of suggested changes to the current IP Version 6 Addressing Architecture”, RFC 2373 [3] to enhance privacy.

A. The Internet Protocol version 6

The Internet Protocol Version 6 (IP$v6$), also called ”IPng” (IP Next Generation), is the latest version of the Internet Protocol (IP). Formally, IPv6 is a set of specifications from the Internet Engineering Task Force (IETF). IPv6 was designed as an evolutionary set of improvements to the current IP Version 4. The most obvious improvement in IPv6 over the IPv4 are that IP addresses are lengthened from 32 bits to 128 bits which anticipates the future growth of the Internet and provides relief for what was perceived as an impending shortage of network addresses. Besides, IPv6 offers technical advantages over IPv4, including self-configuration mechanisms, enhanced security, quality of service features and native mobility support [1].

IPv6 includes a security protocol in the network layer that provides cryptographic security services that supports combinations of authentication, integrity, access control, and confidentiality. The IP Encapsulating Security Payload (ESP) and the IP Authentication Header (AH) are part of the IP Security architecture (IPSEC) described in RFC 2401 [2].

Both ESP and AH are mandatory parts of IPv6 and make sure that a third party eavesdropping on the channel can not read and/or modify the IP datagram. The IP Authentication Header seeks to provide security by adding authentication information to each IP datagram whilst confidentiality requires the use of ESP. Neither AH nor ESP hide the source and destination IP addresses of the communicating parties and hence their network location.

The protocol operation defined for mobility in IPv6 is known as MobileIPv6 and allows a mobile node to move from one link to another without changing the mobile node's IP address. A mobile node is always addressable by its ”home address”, an IP address assigned to the mobile node within its home subnet, i.e., with the prefix of its home link. Packets may be routed to the mobile node using this address regardless of the mobile node's current point of attachment to the Internet, and the mobile node may continue to

communicate with other nodes while using this address, even after moving to a new link. With specific support for mobility in IPv6, packets destined to a mobile node would be able to reach it even while the mobile node is away from its home network.

In summary, IPv6 provides new security opportunities which include message integrity, authentication, and confidentiality (IPSEC) and the possibility for a mobile node to be always addressable by its "home address" (MobileIP). All these functionalities rely on treating the fixed IP address of the node as an identifier. In the case of IPSEC end-to-end security uses the fixed IP address as part of the security association and mobility requires to the mobile node to send the fixed home address included in a destination option.

In the next section, as part of the description of the different mechanisms to obtain a global address, we present the possible threats for privacy when an IP identifier can be linked with personal identifiable information as a "personal device" and how the existing privacy extension has not taken into consideration that the user might be also interested in hiding the fact that is using the privacy extension itself.

II. Mechanisms to obtain a global address

We have classified the mechanisms to obtain a global address as follows: stateless and stateful address configuration, manual, cryptographic generated and random addresses which include the privacy extension and IPv6 over PPP with no global identifier available.

A. Stateless Address Autoconfiguration in IPv6

Stateless address autoconfiguration defines the mechanism for a IP_{v6} node to generate an address without the need for an external DHCP server based on the interface identifier [8,4]. The stateless approach is used when a site is not particularly concerned with the exact addresses hosts use.

To insure that all configured addresses are likely to be unique on given link, nodes do "duplicate address detection" on addresses before assigning them to an interface. The Duplicate Address Detection (DAD) algorithm is used to check all addresses, *independent* of whether they are obtained via stateless or stateful autoconfiguration.

1) Interface Identifier: In the case of stateless address autoconfiguration for Ethernet the Interface Identifier is based on the EUI-64 identifier derived from the interface's built- in 48-bit IEEE 802 address (MAC address).

The EUI-64 is formed as follows:

1) The first three octets that corresponds to the Organizationally Unique Identifier (OUI) of the Ethernet address become the *company_id* of the EUI-64. The OUI blocks are assigned by IEEE.
2) The fourth and fifth octets of the EUI-64 are set to the fixed value FFFE hexadecimal (encapsulation of IEEE 802 address in EUI-64).
3) The last three octets of the Ethernet address (extension identifier values) become the last three octets of the EUI-64.

The Interface Identifier is then formed from the EUI-64 by complementing the "Universal/Local" (U/L) bit, which is the next-to-lowest order bit of the first octet of the EUI-64. Complementing this bit will generally change a 0 value to a 1, since an interface's built-in address is expected to be from a universally administered address space and hence have a **globally unique value**.

A universally administered IEEE 802 address or an EUI-64 is signified by a 0 in the U/L bit position, while a globally unique IPv6 Interface Identifier is signified by a 1 in the corresponding position. In brief, a globally unique IPv6 interface based on a hardware token that is unique will carry one bit indicating its uniqueness.

[3] states that motivation for inverting the "u" bit when forming the interface identifier is to make it easy for system administrators to hand configure local scope identifiers when hardware tokens are not available and to allow development of future technology that can take advantage of interface identifiers with global scope.

2) Link, site and global-scope addresses: The link-local address of an IP_{v6} node is the result of combining the global unique interface identifier with the reserved link-local prefix FE80. The site-local and global-scope addresses are created by combining prefixes advertised in Router Advertisements and Neighbor Discovery.

The IP_{v6} address generated via Stateless Autoconfiguration contains the same interface identifier *regardless* of the location the mobile node is attached to the Internet.

Even when higher communication layers encrypt their payloads (for example with ESP), there is not an easy

mechanism to hide the addresses in packet headers as they appear in clear, this fact makes it very easy for an eavesdropper to track mobile nodes by analyzing the prefixes related to a certain interface identifier.

For example, let's consider a device with built-in 48 bit EUI-48 address 00:01:02:65:71:37 that moves from a network A with prefix 3ffe:200:15:1 to a network B with prefix 3ffe:200:17:2.

1) A unique EUI-64 value is generated by concatenating the company_id, an $FFFE_{16}$ valued label, and the extension identifier values, obtaining 00:01:02:ff:fe:65:71:37.
2) The global unique interface identifier is then formed by complementing the "Universal/Local" bit, resulting 02:01:02:ff:fe:65:71:37.
3) The link local address is fe80::201:2ff:fe65:7137 in both networks
4) When moving from network A to B the mobile device will change from 3ffe:200:15:1:201:2ff:fe65:7137 global IP_{v6} address to 3ffe:200:17:2:201:2ff:fe65:7137.

B. Privacy Extension for Stateless Address Configuration

Narten and Draves [6] developed a privacy extension for Stateless Address Configuration based on the idea of generating random interface identifiers periodically. They describe two approaches for the maintenance of the randomized interface identifier depending upon weather stable storage is present (history scheme) or not. The use of the history scheme tries to avoid the scenario where two nodes generates the same randomized interface identifier, both detect it via DAD, but then proceed to generate identical randomized interface identifier via the same flawed random number generation algorithm.

The first big difference between the EUI-64 based interface identifiers and RFC3041 is that the latest can not claim to be globally unique and hence the universal bit must be set to zero.

C. Stateful address configuration

The dynamic host configuration protocol (DHCP) is the stateful counterpart to stateless autoconfiguration in which hosts obtain interface addresses and/or configuration information and parameters from a server. Servers maintain a database that keeps track of which addresses have been assigned to which hosts. Stateless and stateful autoconfiguration complement each other. For example, a host can use stateless autoconfiguration to configure its own addresses, but use stateful autoconfiguration to obtain other information.

D. Manual configuration

If an IEEE global identifier is not available a different source of uniqueness should be used. Suggested sources of uniqueness include link-layer addresses, machine serial numbers, etc. In this case the "u" bit of the interface identifier must be set to zero.

If a good source of uniqueness cannot be found, it is recommended that a random number be generated. In this case the "u" bit of the interface identifier must also be set to zero.

E. Cryptographically Generated Addresses

The Cryptographically Generated Address where introduced to solve the problem of address ownership [14] in MobileIP. In all the different proposals the CGA addresses have a strong cryptographic binding with a public key. CGA are obtained by means of a one-way hash functions.

For example in the case of SUCV IDs [7] the CGA addresses are created as follows:

$$CGA = 64bit - prefix + CGIID$$
$$CGIID = f(PublicKey, j)$$

where $f()$ is the least significant 64 bits of SHA-1 hash of Public Key concatenated with 16 bit counter j. The universal bit u is set to zero.

F. Classification of Interface Identifiers

The IPv6 interface identifiers can be classified as follows:

1) Half of this space, with the universal bit set to 1, is given to IEEE EUI-64 identifiers.
2) The other half, with the universal bit set 0 is used for Interface Identifiers that are not globally unique including:

- manually assigned.
- local unique assigned by DHCPv6.
- random interface identifiers used in RFC 3041 [6] and RFC 2472 [5] section 4.1.
- cryptographically generated addresses.

III. UNOBSERVABILITY OF THE PRIVACY EXTENSION

In [15] unobservability is defined as the property that ensures that a user may use a resource or service without others, especially third parties, being able to observe that the resource or service is being used.

Unobservability requires that a third party cannot determine whether an operation is being performed based on certain knowledge of the subject(s) of observation.

In our analysis we define as the item of interest for the attacker to be able to determine if the victim is using RFC3041. In order to try to describe different scenarios it is important to identify possible sources of knowledge for the attacker to make a decision. The basic source of information is the link local or global scope address of the victim but other sources could be available as:

- The victim's mac address.
- The presence of a DHCPv6 serving the prefix of the victim's network.
- If the victim has mobility support and if its address is a CGA.
- The victim's hardware type and/or operative system.

IV. SCENARIOS

According to the amount of knowledge available to the attacker we can divide all scenarios in two.

A. Attacker is not present in the link

In this scenario the attacker is not present in the link and is in the path between the victim and a correspondent node. In our model we assume that the attacker knows about which OUIs has been assigned by IEEE but not the distribution of the "extension identifier values" for certain OUI.

In first place the attacker determines if the interface identifier is based on a EUI-64 identifier by checking the universal/local bit. If the universal/local bit is set to one, the attacker checks if the OUI has been assigned by IEEE. If the OUI has not been assigned by IEEE the attacker can presume that the victim has either modified its MAC address, configured manually the interface identifier or generated a random interface identifier without setting the u-bit to zero.

If the universal/bit is set to zero, the attacker can choose among different possibilities : the interface

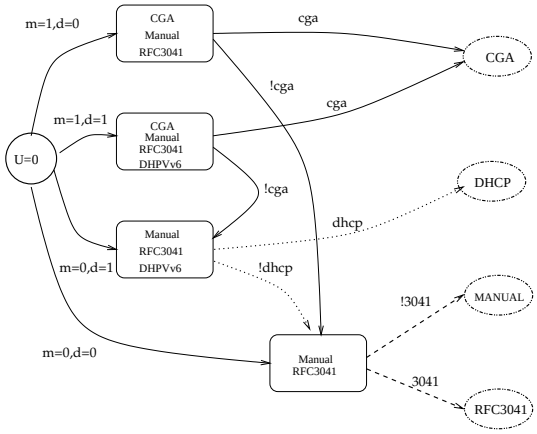

Fig. 1. Possible scenarios

has been configured by stateful address configuration, manualy or is random interface identifier (CGA, RFC3041, RFC2472).

The attacker can gather extra knowledge as follows:

- By observing the nature of the traffic between the victim and a correspondent node. Cryptographically Generated Addresses are included in authentified binding updates and are part of destination/routing header options.

- By discovering the presence of a DHCP server via the (FF05::1:3) site-scoped multicast address.

B. Attacker is present in the link

In the second main scenario the attacker is present in the link, most of the information required to observe if the victim is using the privacy extension can be gathered by victim's traffic analysis. The attacker will be able to observe not only the presence of a DHCP server but also the traffic exchange between the DHCP server and the victim. If the victim has a CGA, the attacker will be also able to observe the flow of authentified mobility bindings.

In the case that the victim is not using stateful address configuration or is a mobile node with a CGA, the attacker has to decide between two possible options [Fig. 1]: the victim is running RFC3041 privacy extension or has configured manually its address. The attacker then, can try to match during a certain period of time the hardware addresses with the link or global scope addresses. Nodes running RFC3041 will change their addresses periodically while keeping the same hardware address.

C. Conclusions

[Fig. 1] shows all possible scenarios considering the amount of knowledge available to the attacker. In all the cases the attacker starts checking if the universal/local bit of the interface identifier is set to zero. If the node is not running MobileIP with CGA and there is not a DHCP server available in the victim's subnet, the attacker assumes that the victim is running RFC3041 or has configured the address manually. Finally, the attacker can observe the addresses associated with a certain hardware address and determine if the victim is running RFC3041.

V. Recommendations

A better privacy protection can be achieved if the random interface identifier can not be distinguished from a *common* one. i.e. an eavesdropper can not determine if certain node is using or not the stateless address configuration privacy extension.

Unobservability can be guaranteed as follows:

- All the hosts generate their interface identifier randomly by default. (suggested change in RFC2373).

- The universal/local bit is not reserved and hosts always rely in duplicate address detection (DaD).

- Alternatively, the host generates an interface identifier based on the addresses present in the link. The main idea is that the mobile node should keep statistical records of the presence of the different OUIs in the media and generate a random identifier based on that information. The host learns about the nodes in the media by sending a neighbor discovery message to the all hosts multicast address.

References

[1] **C. Huitema,** *IPv6, the new Internet Protocol.* 2nd Edition. Prentice Hall. 1997

[2] **S. Kent and R. Atkinson**, *"Security Architecture for the Internet Protocol"*. RFC 2041.

[3] **R. Hinden and S. Deering**, *"IPv6 Addressing Architecture"*, RFC 2373, July 1998.

[4] **R. Hinden, M. O'Dell, S. Deering**, *"An IPv6 Aggregatable Global Unicast Address Format"*, RFC 2374, July 1998.

[5] **D. Haskin and H. Allen** *"IP Version 6 over PPP"*, RFC 2472, December 1998

[6] **T. Narten and R. Draves**, *"Privacy Extensions for Stateless Address Autoconfiguration in IPv6"*. RFC3041, January 2001.

[7] **G. Montenegro and C. Castelluccia,** *"SUCV Identifiers and Addresses"*. draft-montenegro-sucv-01.txt. July 2001.

[8] **S. Thomson and T. Narten,** *"IPv6 Address Autoconfiguration"*, RFC 2462, December 1998.

[9] **A. Escudero. "***Location privacy in IP_{v6} internetworking - Pseudorandom interface identifiers"*. IDMS2001. Lancaster. UK. August 2001.

[10] **D. Johnson and C. Perkins,** *"Mobility Support in IPv6"*, draft-ietf-mobileip-ipv6 v14, July 2000.

[11] **H. Soliman, C. Castelluccia, K. El-Malki and L. Bellier,** *"Hierarchical MIPv6 mobility management"*, draft-ietf-mobileip-hmipv6 v3. February 2001.

[12] **C. Perkins,** *"IP Mobility Support"*, RFC 2002, October 1996.

[13] **A. Escudero**, *"Anonymous and Untraceable Communications: Location Privacy in Mobile Internetworking"*. Licentiate Thesis. July 2001.

[14] **Pekka Nikander**, *"An Address Ownership Problem in IPv6"*, draft-nikander-ipng-address-ownership-00.txt, February 2001.

[15] **ISO99** IS 15408, 1999, `http://www.csrc.nist.gov/cc`

PAPER #4

Alberto Escudero-Pascual

"Privacy enhanced architecture for location based services in the next generation wireless networks"

11th IEEE Workshop on Local and Metropolitan Area Networks (LANMAN2002), pp. 169-172.
Stockholm, Sweden
August 2002

PRIVACY ENHANCED ARCHITECTURE FOR LOCATION BASED SERVICES IN THE NEXT GENERATION WIRELESS NETWORK

Alberto Escudero-Pascual

IMIT, Royal Institute of Technology
Isafjorsgatan 39, Stockholm, Sweden, <aep@kth.se>

Abstract - Location-based services (LBS) can be described as applications that exploit knowledge about where an information device (user) is located. For example, location information can be used to provide automobile drivers with optimal routes to a geographical destination or a group of friends with the names and coordinates of Spanish's restaurants in the neighborhood open on a Saturday night.

We propose a privacy enhanced location based service (PE-LBS) architecture which allows a mobile node to request location based services via a proxy server hiding the network location of the mobile device while providing service accountability.

The architecture is composed of six modules: location acquisition hardware, XML data record parser [1], XML service request, transport module, LBS proxy and service modules. Privacy Enhanced Technologies has been carefully integrated to enhance the privacy of our architecture by protection of personal identifiable information.

One of the components of our architecture is the LBS Proxy Server, responsible of processing SOAP [2] (message envelope) requests and generate responses. When the SOAP request is received by a server, it gets bound to the class specified in the request. The proxy server works as a SOAP Dispatcher, by determining which class should handle a given request, and loading that class, if necessary. The SOAP server acts as an intermediary between a SOAP client and the requested service provider.

Keywords - privacy, location based services, privacy enhanced technologies.

I. LOCATION BASED SERVICES ARCHITECTURE

The proposed privacy enhanced location based service (PE-LBS) architecture is composed of six functional modules as follows:

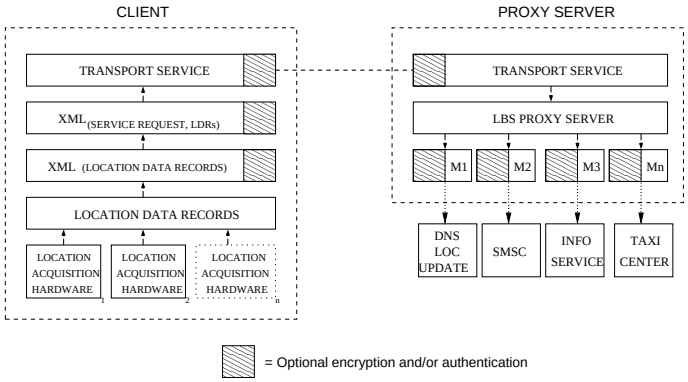

Fig. 1. PE-LBS Architecture

A. Location Acquisition Hardware

The location acquisition hardware is responsible for calculating the position of the mobile device based on a set of data inputs that can vary from GPS radio signals or infrared beacons to an enhanced tape measure. The output is a set of coordinates based on a reference system. For example, most GPS receivers use a global reference system named WGS 84 (World Geodetic System 1984).

Location information records obtained from the hardware can include: latitude, longitude, altitude, velocity, horizontal error, vertical error, global error, orientation, etc.

B. XML Location Data Record

The format of the location records provided by the hardware or multiple pieces of hardware can be of very different nature. The XML Location Data Record module is responsible for creating XML output based on location information provided by the location acquisition hardware.

C. XML Service Request

The XML service request module will take the location information from the XML location data record

and build a service request. In our architecture the service request uses Simple Object Access Protocol (SOAP) [2] to encapsulate and exchange RPC calls using the extensibility and flexibility of XML. SOAP can potentially be used in combination with a variety of other protocols; however, the most common use of SOAP is in combination with HTTP, the experimental HTTP Extension Framework, or SNMP.

D. Transport Service

This module implements the equivalent of OSI layer 4 by providing reliable transparent data transfer between end points, along with error recovery, and flow control. It is responsible for the transport of the remote procedure call to the location based service proxy server.

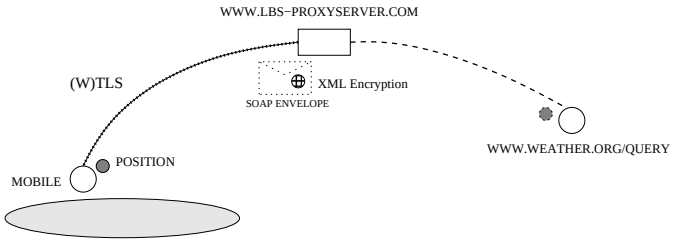

Fig. 2. SOAP Request via PE-LBS proxy

E. Location Based Service Proxy Server

The functionality of the LBS Proxy Server is to process SOAP (message envelope) requests and generate responses. When the SOAP request is received by a server, it gets bound to the class specified in the request. The proxy server works as a *SOAP Dispatcher*, by determining which class should handle a given request, and loading that class, if necessary. The SOAP server acts as an intermediary between a SOAP client and the requested service provider.

F. Service Modules

A service module acts as a SOAP interface, a frontend that requests information or the execution of a procedure, parses and formats the response and returns it according to the request (if necessary). The procedure can run in the same server (e.g. return a prime number of n bits) or be the result of a call in a remote server (e.g. send an e-mail message to a certain address).

Let us consider the scenario where a mobile device with a unique identifier *mobileID* requests the temperature information for a certain position and time. In this case, the SOAP server (LBS-proxy) works as a proxy of the SOAP client (mobile device) and the temperature service provider. In fact, the proxy can conceal from the temperature server the mobile device's *mobileID* and protect its identity as this information is not required to obtain the requested service. But, must the proxy know about the location of the mobile device to proxy the temperature request service? No, we can hide the position information from the proxy and still get the temperature in that position. To do this we use a privacy enhanced proxy.

II. MIXES AND PE-LBS PROXIES

David Chaum described in [3] a technique based on public key cryptography that allows an electronic mail system to hide who a participant communicates with as well as the content of the communication.

More generally, messages are exchanged through a chain of one or more intermediaries called ”mixes”. The purpose of a mix is to hide the correspondences between the items in its input and those in its output. The main function of a mix is to: receive and decrypt messages, buffer messages until a defined number of messages has been received, change the sequence of the received messages in a random manner and encrypt and forward the messages to the next mix or to the receiver.

Three of the benefits of our architecture are: the possibility of a PE-LBS proxy to act as a ”mix” by buffering and changing the sequence of the service requests, a mobile device can use a chain of PE-LBS proxies configured as a ”mixing network” to forward a location based service requests and that these functionalities can be done independently of the specific transport network.

III. A PROOF OF CONCEPT

A proof of concept was implemented using Fastrax's iTrax02 GPS receiver. The iTrax02 is an ultra-low power consumption receiver, roughly the size of a stamp and specifically designed for small portable devices. In one of the scenarios, the location information is encrypted using a public key encryption scheme (with multiple private keys), embedded in a XML message and transmitted to a proxy that runs a secure DNS update module [5]. This location privacy solution allows a mobile terminal to publish its location as an encrypted DNS location record via the proxy, while

concealing from eavesdroppers and third parties the relation between the location information and the identity of the mobile terminal and its user.

IV. Conclusions

By using a proxy server between the mobile node and the location based service we have shown that we can hide the network location of the mobile device and in some cases even provide misleading physical location(s) for the mobile device [4].

Combining XML Encryption with XML Signature in Simple Object Access Protocol service requests provide both message digest and message authentication functionality. Taking advantage of the extensibility and flexibility of XML it is easy to implement and extend the set of privacy enhanced location based services while still hiding the mobile node's network and physical location as desired[1].

V. References

[1] W3C "XML Encryption Syntax and Processing", Working Draft, 18 October 2001.

[2] W3C "Simple Object Access Protocol (SOAP) 1.1", Technical Report. May 2000.

[3] D. Chaum. "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms". Communications of the ACM (24)2, 1981 pp. 84-88

[4] A. Escudero. "Privacy and Identity in the Information Society - Protection of personal identifiable information in mobile Internet". IPSC-IPTS (EU). Brussels. October 2001.

[5] C. Davis, et al. "RFC 1876: A Means for Expressing Location Information in the Domain Name System", January 1996.

- - -

[1] This work would not be possible without the support and advice of Gerald Q. Maguire Jr.

PAPER #5

Alberto Escudero-Pascual and Gerald Q. Maguire Jr.

"Role(s) of a proxy in location based services"

13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC2002), Vol.3 pp. 1252-1257
Lisbone. Portugal
September 2002

ROLE(S) OF A PROXY IN LOCATION BASED SERVICES

Alberto Escudero-Pascual[1], Gerald Q. Maguire Jr.[2]

[1] IMIT, Royal Institute of Technology, Isafjorsgatan 39, Stockholm, Sweden, aep@kth.se
[2] Wireless@KTH, Royal Institute of Technology, Stockholm, Sweden, maguire@kth.se

Abstract - We examine a number of roles that a proxy server can play in Location Based Services and how it can be used to provide protection of personal identifiable information. Location data, service requests, and privacy policies are encoded in XML by the mobile terminal and forwarded to a proxy server placed between the mobile terminal and the location based service(s). We will show that by a suitable architecture in the mobile terminal and in the proxy that we can hide the network location of the mobile device, hide the identity of the user of the mobile device, and in some cases even provide misleading physical location(s) for the mobile device. We will illustrate a number of different functions which can be provided by examining some scenarios.

In order to illustrate our approach, we have applied our privacy model to location information obtained from a Global Positioning System receiver. Among the different methods to obtain a mobile's position the GPS-based method was chosen as being the only method, available today, where the Positioning Calculation Function (PCF) is fully under the user's control, since the position is calculated within the GPS-equipped mobile terminal; while other technologies rely on the network infrastructure and hence some or all of the position data is outside the control of the user.

A proof of concept was implemented using Fastrax's iTrax02 GPS receiver. The iTrax02 is an ultra-low power consumption receiver, roughly the size of a stamp and specifically designed for small portable devices. In one of the scenarios, the location information is encrypted using a public key encryption scheme (with multiple private keys), embedded in a XML message and transmitted to a proxy that runs a secure DNS update module. This location privacy solution allows a mobile terminal to publish its location as an encrypted DNS location record via the proxy, while concealing from eavesdroppers and third parties the relation between the location information and the identity of the mobile terminal and its user.

Keywords - privacy, location based services, privacy enhanced technologies.

I. INTRODUCTION

Location-based services (LBS) can be described as applications that exploit knowledge about where an information device (user) is located. For example, location information can be used to provide automobile drivers with optimal routes to a geographical destination or a group of friends with the names and coordinates of Spanish's restaurants in the neighborhood open on a Saturday night.

Location information can be used as external input for applications, but can also be used by lower layers in combination with link level information in mobile networks to optimize network performance, for example in assisting in handover decisions for MobileIP [1].

When talking about cellular networks, location-based services exploit any of several technologies for determining where a network user is geographically located. ETSI has issued a specification [2] that deals with different methodologies to obtain location information of a mobile station as follows: Time Advance (TA), Time of Arrival (TOA), Enhanced Observe Time Difference (E-OTD), Angle of Arrival (AOA) and Global Positioning System (GPS).

Depending on where the information is gathered and the position calculation function (PCF) is computed the different methods to obtain the position can be divided in four categories: network based, network based-mobile assisted, network assisted-mobile based, and mobile based [2].

This paper deals with the situations when then the position is computed in the terminal (i.e., mobile based) with or without assistance from the network. In these scenarios the user is in control of the location information associated with the mobile device.

However, problems arise when the user needs to provide that information in order to obtain a service and at the same time doesn't want to reveal more personal identifiable information that is strictly necessary [3]. For example, a mobile user may want to inform to only a certain number of people for a certain period of time about his or her position or, to learn the position of the nearest catholic church without revealing his or her personal identity.

The paper is divided as follows:

SECTION 2 contains the description of our Location Based Services Architecture and the role of the proxy server.

SECTION 3 describes some technologies that can be integrated in some of the modules and will empower users to control their personal information.

SECTION 4 introduces GPS positioning, NMEA messages, and the iTrax02 GPS receiver.

SECTION 5 illustrates a number of different services which can be provided by examining some scenarios.

This is followed by conclusions, the bibliography and an appendix that includes some of the code used to illustrate the architecture.

II. LOCATION BASED SERVICES ARCHITECTURE

We propose a privacy enhanced location based service (PE-LBS) architecture composed of six functional modules which allows a mobile node to request location based services via a proxy server. Once the basic functionalities are described, we will show that by using our architecture in the mobile client and in the proxy that we can hide the network location of the mobile device and hide the identity of the user of the mobile device.

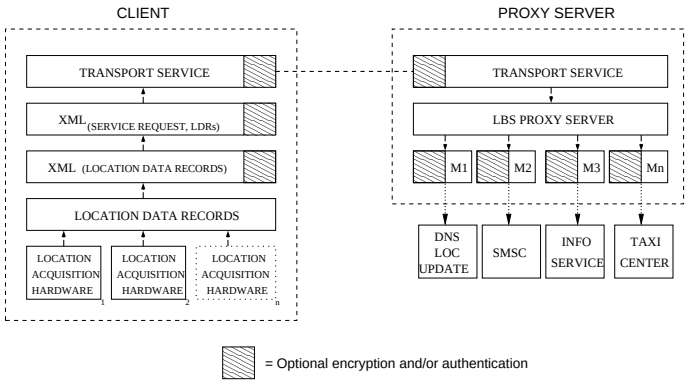

Fig. 1. PE-LBS Architecture

A. Location Acquisition Hardware

The location acquisition hardware is responsible for calculating the position of the mobile device based on a set of data inputs that can vary from GPS radio signals or infrared beacons to an enhanced tape measure. The output is a set of coordinates based on a reference system. For example, most GPS receivers use a global reference system named WGS 84 (World Geodetic System 1984).

Location information records obtained from the hardware can include: latitude, longitude, altitude, velocity, horizontal error, vertical error, global error, orientation, etc.

B. XML Location Data Record

The format of the location records provided by the hardware or multiple pieces of hardware can be of very different nature. The XML Location Data Record module is responsible for creating XML output based on location information provided by the location acquisition hardware. An example of a XML location data output encoded based on [4,5] looks like [Appendix: code1].

C. XML Service Request

The XML service request module will take the location information from the XML location data record and build a service request. In our architecture the service request uses Simple Object Access Protocol (SOAP) [7] to encapsulate and exchange RPC calls using the extensibility and flexibility of XML. SOAP can potentially be used in combination with a variety of other protocols; however, the most common use of SOAP is in combination with HTTP, the experimental HTTP Extension Framework, or SNMP.

The example included in [Appendix: code2] can be translated as: Ask http://www.lbs-proxyserver.com (a SOAP server) to provide the temperature (GetTemperature) at latitude: N59.40.54 and longitude: E017.94.36 at 2001-01-01T12:00:01+02:00 from http://weather.org/query

D. Transport Service

This module implements the equivalent of OSI layer 4 by providing reliable transparent data transfer between end points, along with error recovery, and flow control. It is responsible for the transport of the remote procedure call to the location based service proxy server.

E. Location Based Service Proxy Server

The functionality of the LBS Proxy Server is to process SOAP (message envelope) requests and generate responses. When the SOAP request is recieved by a server, it gets bound to the class specified in the request. The proxy server works as a *SOAP Dispatcher*, by determining which class should handle a given request, and loading that class, if necessary. The SOAP server acts as an intermediary between a SOAP client and the requested service provider.

F. Service Modules

A service module acts as SOAP interface, a frontend that requests information or the execution of a procedure, parses and formats the response and returns it according to the request (if necessary). The procedure can run in the same server (e.g. return a prime number of *n* bits) or be the result of a call in a remote server (e.g. send an e-mail message to a certain address).

Let us consider again the example presented in [Appendix: code1], where a mobile device with a unique identifier *mobileID* requests the temperature information for a certain position and time. In this case, the SOAP server (LBS-proxy) works as a proxy of the SOAP client (mobile device) and the temperature service provider. In fact, the proxy can conceal from the temperature server the mobile device's *mobileID* and protect its identity as this information is not required to obtain the requested service. But, must the proxy know about the location of the mobile device to proxy the temperature request service? No, we can hide the position information from the proxy and still get the temperature in that position. To do this we use a privacy enhanced proxy.

III. Privacy enhanced LBS architecture. *PE-LBS*

The architecture described below is composed of six modules: location acquisition hardware, XML data record parser, XML service request, transport module, LBS proxy and service modules. In this section we are going to describe some technologies that carefully integrated can enhance the privacy of our architecture by protection personal identifiable information.

A. XML encryption

XML Encryption [6] is a recently developed cryptographic format that describes the process for encrypting data and representing the result in an XML Encryption element which contains or identifies the cipher data. XML Encryption may be used on a whole XML document, an XML entity, an XML entity content, or on arbitrary binary data.

XML Encryption supports both symmetric cryptographic algorithms (AES and Triple DES) and asymmetric cryptographic algorithms (also referred to as key transport algorithms such as RSA).

After encryption the resulting cipher text is either included within the original XML document encoded as a base64 octet sequence or if the cipher text is located outside the document an URI reference reveals the location where the cipher text can be found. The <EncryptedData> entity replaces the original entity or entity content. The <EncryptedData> entity contains both the cipher text and related information, needed for decryption of the cipher text into plain text.

In summary, by combining XML Encryption with XML Signature we can provide both message digest and message authentication functionality.

Consider again the example presented in [Appendix: code1], by using XML encryption we can conceal from the proxy server the values LAT, LONG, TIME of the GetTemperature request. The example [Appendix: code3] shows the use of *3des-cbc* symmetric encryption to encrypt the content of the GetTemperature request.

B. Transport Security Protocols

A transport security protocol provides confidentiality; data integrity, and authentication for information exchanged between a client and a server from the session layer and above in the OSI model and in the WAP stack. The most frequent used protocol for securing plain text transmissions for wired usage is IETF's Transport Layer Security (TLS) based on Netscape's Secure Socket Layer (SSL) [9].

For wireless usage the WAP forum developed the optional Wireless Transport Layer Security (WTLS) protocol that operates below the WDP and WTP protocols [8].

A transport security protocol as (W)TLS can be used to provide confidentiality; data integrity, and authentication for SOAP messages exchanged between the LBS proxy server and the mobile device.

IV. Private LBS with GPS-equipped mobile terminal

A. Introduction to GPS positioning

The Global Positioning System (GPS), originally a US military technology, was made available for civilian uses by its owner the US Department of Defense in the early 1990's. The GPS system is one of the most accurate navigation systems available today. The system consists of a network of 24 active satellites orbiting the Earth once every 12 hours and in six orbits orbits at inclinations of 55 degrees from the equator and at aproximately 20200 Km altitude.

Each satellite is carrying atomic clocks for transmitting timing signals worldwide. Any observer who can receive signals from four of these satellites can determine his accurate position on earth. The satellites orbit the earth twice each day which allows a receiver to see from five to eight of them from any position on the Earth at anytime.

To calculate its position, the receiver needs to know at least the position of three satellites (four satellites are needed to include the time) and the distance from the GPS receiver to the satellites.

In order to know where the satellites are located the GPS receiver collects the almanac and ephemeris information from the radio signals. The "almanac" contains information about the satellite's orbit and provides the approximate location of the satellite, the second type of information, "ephemeris", is uploaded to the satellites from groundstations and provides for a period of four to six hours the neccesary corrections to adjust the planned orbit information to the actual orbit.

The receiver measures the time required for the signal to travel from the satellite to the receiver, by knowing the time that the signal left the satellite, and observing the time it receives the signal, based on its internal clock.

If the receiver had a perfect clock, exactly in synchronization with those on the satellites, three measurements, from three satellites, would be sufficient to determine the receiver's position in 3 dimensions. Each measurement ("pseudorange") gives a position on the surface of a sphere centred on the corresponding satellite. Due to receiver clock error, the four spheres will not intersect at a single point, but the receiver will adjust its clock until they do, providing very accurate time, as well as position information. Since the receiver must adjust its clock to be precisely in synchronization with GPS time, a GPS receiver can be used as a precise time reference.

B. NMEA General Message Format

The National Marine Electronics Association (NMEA) issues standards for interfacing to marine electronics. NMEA 0183 is a standard protocol, use by GPS receivers to transmit data to attached devices. The data transmition accours at 4800 bps, with 8 data bits, no parity, and one stop bit (8N1).

In the case of GPS data the messages starts with '$GP' followed by message id field. Message data fields are separated by commas and the message ends after the checksum field and carriage return and line feed control characters. Delimiter $'*'$ precedes the checksum field.

For example the NMEA 0183 message: $\$GPGLL, 5924.3131, N, 01756.5752, E, 134703.77, A, A*61$ is a Geographic Position Latitude/Longitude message (GLL) that provides: Latitude, Longitude, UTC time of fix and status.

C. About iTrax02

Fastrax's GPS (Global Positioning System) receiver, shown in Fig. 2, is roughly the size of a postage stamp, 25x25x4 mm. The small foot print combined with ultra-low power consumption (130mW in full operational mode) and low cost make it feasible to utilize GPS positioning technology in mass-market applications, particularly those designed for small portable devices, in which low power consumption and small size are crucial parameters, such as in mobile phones, sports instruments, and handheld computers.

Fastrax receivers have very good sensitivity. Receiver sensitivity is one of the most important features of all GPS receivers as all other functions are dependent on receiver's ability to find satellites fast and receive the information from the satellites and convert these measurements into a 3D navigational vector. iTrax02 is able to produce and interpret standard NMEA as well as their own binary format *iTalk*.

V. Scenarios

In this section we are going to describe some scenarios where location information is required to access/provide a service and how our privacy enhanced location based services architecture can be introduced in each of them.

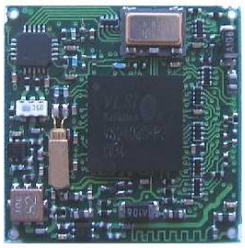

Fig. 2. iTrax02: 25x25x4 mm GPS receiver

A. Publishing location information in DNS LOC resource records

1) Introduction to DNS Resource Record LOC and TSIG: The DNS LOC is described in [10], DNS LOC is a new Domain Name Server Resource Record (DNS RR) type for experimental purposes and describes a mechanism to allow the DNS to carry location information about hosts, networks, and subnets. The records contains information about latitude, longitude, altitude, horizontal and vertical error, and size of the described entity.

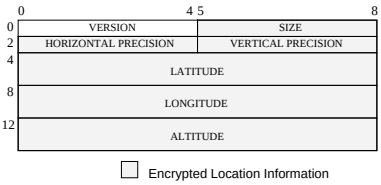

Fig. 3. DNS LOC. RDATA

The Transaction signatures (TSIG) provide an authentication mechanism that uses shared secret keys to establish a trust relationship between two entities. TSIGs are described at [11] along with the way that DNS messages should be treated by a forwarding server. If the name on the TSIG is not of a secret that the server shares with the originator the server must forward the message unchanged including the TSIG.

2) Description of the scenario : In this scenario a mobile node wants to make available its location only to a set of correspondent nodes. The mobile node shares a set of secrets with the DNS server and the correspondent nodes.

- **Location Acquisition Hardware:** At a certain time the iTrax02 GPS receiver sends a GPGLL NMEA 0183 message with the location information.
- **XML data record:** The location and time information is extracted from the NMEA message and converted to XML format.

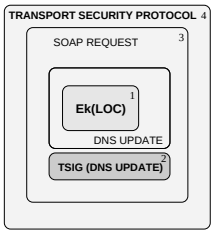

Fig. 4. SOAP request (DNSSec LOC Update)

- **XML service request:** A SOAP request is created as follows:
  - Instead of including the original location data in the DNS LOC update, the LOC RDATA is encrypted with the secret shared with the correspondent nodes. A DNS Update message is created that contains the encrypted location information, shown in Fig. 4 as item 1.
  - Once the outgoing DNS update message has been constructed, the keyed message digest operation can be performed *(hmac-md5)* using the shared secret with the DNS server. The resulting digest message will then be stored in a TSIG which is appended to the additional data section, shown in Fig. 4 as item 2.
  - The DNSUpdate service request can be contructed as a SOAP request to the LBS proxy server. In the body a TSIG DNS Update message is included as an entity in the request <DNSSecUpdate>, shown in Fig. 4 as item 3.
- **Transport Security Protocol:** A Secure Socket Layer is established between the mobile node and the LBS proxy. The SOAP request is used as a method invocation mechanism encapsulated in a HTTP over the SSL tansport channel, shown in Fig. 4 as item 4.
- **LBS Proxy Server:** When the SOAP request is recieved by the LBS server, it gets bound to the DNSUpdate method specified in the request.
- **DNS Update Service Module:** The DNS Update module extracts the DNS Update message from the XML and sends a DNS Update message to the remote DNS server. Thus the DNS Update Service Module acts as a TSIG DNS forwarding server.

Note that (1) the DNS Server, accepts the DNS update based on the TSIG, (2) neither the proxy nor the DNS server knows the mobile node coordinates - since they can't read the encrypted LOC RDATA, and (3)

the DNS server can't even know the mobile's current network attachment point. The mobile's location is now available to any node which knows the key to decrypt the results of a DNS LOC query.

B. Restaurant Info request

Let us consider another scenario when a mobile node wants to request the GPS coordinates of the spanish restaurants in the neighborhood open Saturday night, in this case the location information and the nature of the information requested (spanish restaurant, open Saturday night) is encrypted using a shared secret with the Information Server. The proxy server is only aware that an *Info request* has been submitted, but not the type of information requested (which includes location and time).

The architecture also allows a mobile node to hide its true *location of interest* by including more than one location based request in the same SOAP Body. Sending multiple requests obscures the *location of interest* from the Information Server. The Information Server can't tell if the mobile is actually at or near any of these locations.

Note in this restaurant example, the proxy will see the Information Server's reply/replies, hence the need to either (1) obscure via multiple query results or (2) the mobile must provide a symmetric key to be used by the Information Server to encrypt its reply so the proxy only has to pass on the opaque reply. However, this second case requires the server to implement an additional privacy enhancement.

VI. MIXES AND PE-LBS PROXIES

David Chaum described in [12] a technique based on public key cryptography that allows an electronic mail system to hide who a participant communicates with as well as the content of the communication.

More generaly, messages are exchanged through a chain of one or more intermediaries called "mixes". The purpose of a mix is to hide the correspondences between the items in its input and those in its output. The main function of a mix is to: receive and decrypt messages, buffer messages until a defined number of messages has been received, change the sequence of the received messages in a random manner and encrypt and forward the messages to the next mix or to the receiver.

Three of the benefits of our arquitecture are: the possibility of a PE-LBS proxy to act as a "mix" by buffering and changing the sequence of the service requests, a mobile device can use a chain of PE-LBS proxies configured as a "mixing network" to forward a location based service requests and that these functionalities can be done independently of the specific transport network.

VII. CONCLUSIONS

By using a proxy server between the mobile node and the location based service we have shown that we can hide the network location of the mobile device and in some cases even provide misleading physical location(s) for the mobile device.

Combining XML Encryption with XML Signature in Simple Object Access Protocol service requests provide both message digest and message authentication functionality. Taking advantage of the extensibility and flexibility of XML it is easy to implement and extend the set of privacy enhanced location based services while still hiding the mobile node's network and physical location as desired.

REFERENCES

[1] **C. Perkins,** "RFC2002: IP Mobility Support", October 1996.

[2] **ETSI GSM 03.71**, "Digital cellular telecommunications system (Phase 2+); Location Services (LCS)", 2000

[3] **A. Escudero,** "Protection of personal identifiable information in mobile internet". IPSC-IPTS Workshop. Brussels. October 2001.

[4] **M. Korkea-aho and H. Tang,** "A Common Data Set and Framework for Representing Spatial Location Information in the Internet - Internet Draft", May 2001.

[5] **F. Yergeau,** "RFC 2279: UTF-8, a transformation format of ISO 10646", 1998.

[6] **W3C,** "XML Encryption Syntax and Processing", Working Draft, 18 October 2001.

[7] **W3C,** "Simple Object Access Protocol (SOAP) 1.1", Technical Report. May 2000.

[8] **WAP Forum,** "WAP Wireless Transport Security Specification", February 2000.

[9] **T. Dierks and C. Allen,** "RFC2246: The TLS Protocol Version 1.0". January, 1999.

[10] **C. Davis et al,** "RFC 1876: A Means for Expressing Location Information in the Domain Name System", January 1996.

[11] **P. Vixie et al,** “RFC 2845: Secret Key Transaction Authentication for DNS (TSIG)”, May 2000.

[12] **D. Chaum,** “Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms”, Communications of the ACM (24)2, pp. 84-88, 1981.

```
--- [Appendix: code1] - ’XML Location Data Record’
<?xml version = "1.0" encoding = "UTF-8"?>
<loc:SLO xmlns:loc="http://www-nrc.nokia.com/ietf-spatial/2001/05/08/location"
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
         xsi:schemaLocation="http://www-nrc.nokia.com/ietf-spatial/2001/05/08/location
         http://www-nrc.nokia.com/ietf- spatial/2001/05/08/location.xsd">
  <POS>
    <LAT>N59.40.54</LAT>
    <LONG>E017.94.36</LONG>
  </POS>
  <ALT>+12.99</ALT><H_ACC>50</H_ACC><V_ACC>2.5</V_ACC>
  <TIME>2001-13-11T12:00:01+02:00</TIME>
</loc:SLO>
--- [Appendix: code2] - ’SOAP Service Request’
POST /Temperature HTTP/1.1
Host: www.lbs-proxyserver.com
Content-Type: text/xml
Content-Length: 357 SOAPAction: "http://weather.org/query#GetTemperature"
       <SOAP-ENV:Envelope
          xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
          SOAP-ENV:encodingStyle="http://schemas.xmlsoap.org/soap/encoding/">
         <SOAP-ENV:Body>
          <m:GetTemperature xmlns:m="http://weather.org/query">
            <TIME>2001-01-01T12:00:01+02:00</TIME>
            <LAT>N59.40.54</LAT>
            <LONG>E017.94.36</LONG>
          </m:GetTemperature>
         </SOAP-ENV:Body>
       </SOAP-ENV:Envelope>
--- [Appendix: code3] ’Privacy Enhanced SOAP Service Request’
POST /Temperature HTTP/1.1
Host: www.lbs-proxyserver.com
Content-Type: text/xml
Content-Length: 357 SOAPAction: "http://weather.org/query#GetTemperature"
       <SOAP-ENV:Envelope
          xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
          SOAP-ENV:encodingStyle="http://schemas.xmlsoap.org/soap/encoding/">
         <SOAP-ENV:Body>
          <m:GetTemperature xmlns:m="http://weather.org/query">
           <EncryptedData xmlns=’http://www.w3.org/2001/04/xmlenc#’
              Type=’http://www.w3.org/2001/04/xmlenc#Content’>
           <EncryptionMethod Algorithm=’http://www.w3.org/2001/04/xmlenc#3des-cbc’/>
           <ds:KeyInfo xmlns:ds=’http://www.w3.org/2000/09/xmldsig#’>
           <ds:KeyName>KeyID
            </ds:KeyName>
            </ds:KeyInfo>
           <CipherData>
               <CipherValue>XkIHMHS4ka4CXFWA3yESBqQzIp21D1MHYPG
                            kL7bXoC8S9tQlIKbghAkHbZDgrzBI6yvP33</CipherValue>
           </CipherData>
         </m:GetTemperature>
         </SOAP-ENV:Body>
       </SOAP-ENV:Envelope>
```

PAPER/JOURNAL #6

Alberto Escudero-Pascual and Ian Hosein.

"The hazards of technology-neutral policy: questioning lawful access to traffic data"

To appear in Communications of Association for Computing Machinery Journal
Accepted 5th September 2002 - Reviewed 19th October 2002

THE HAZARDS OF TECHNOLOGY-NEUTRAL POLICY: QUESTIONING LAWFUL ACCESS TO TRAFFIC DATA

.

Policies are being updated to deal with new communications infrastructures;
the path to policy renewal is fraught with danger

.

By Alberto Escudero-Pascual and Ian Hosein

To appear in Communications of ACM (Accepted 5-9-2002, Reviewed 19-10-2002)

Abstract - After some successes and many mis-steps, the regulatory environment surrounding technology policy is transforming. Lessons taken from content, copyright, and cryptography policy processes, amongst many others, resulted in the emergence of a number of technology policy innovations. Two particular innovations are the internationalization of policy-setting, and the trend towards technology-neutral policies. These innovations come with risks, however. The risks are particularly apparent when we look at policies on law enforcement access to traffic data.

Access to traffic data for law enforcement purposes is a traditional tool for investigation and intelligence gathering. *Traffic data* is an elusive term, due in part to technology variances. The policies regarding lawful access to traffic data, however, are increasingly set in technology-neutral language, while the language is often negotiated at international fora.

Each policy innovation needs to be questioned. The momentum behind the policy changes comes from both the technology and international incentive schemes. Yet the policies tend to ignore the technological details; while policy changes are argued as necessary due to international obligations. In the hope of updating our policies, we may be numbing our technological awareness and political openness.

LAW ENFORCEMENT REQUIREMENTS: THE POLITICAL

In the days of plain old telephone systems (POTS), after much legal debate, the content of communications were considered sensitive and therefore any breach of confidentiality, i.e. wiretapping required constraint, e.g. judicial warrants in the U.S., politician-authorized warrants in the United Kingdom. The same rule did not apply to *traffic data*: numbers called, calling numbers, and time. This data was considered less invasive, and therefore only required minimal constraint. That traffic data was stored by telephone companies and thus available to law enforcement authorities while communications were not, also reduced the obstacles: traffic data was available, legally less sensitive, and so accessible. This is the policy habitat [7] of traditional surveillance of communications.

The traffic data records collected by telephone companies generally look like:

[See Appendix I.A]

While the format of the logs may differ from one operator to another, the above log can identify the time and duration of a call, the phone-ID numbers involved in the call, the countries involved, and the types of service used.

The traditional investigative powers of access to traffic data were established with traditional technological environments in mind. Governments are updating their policies on interception of communications to apply to modern communications infrastructures. As cryptography policies of key escrow were mis-understood as updates 'to maintain the status quo' of government powers [9], updating legal definitions of traffic data while not acknowledging the increased 'sensitivity' of the data by claiming technological neutrality is equally problematic.

The claim of technological neutrality

Similar to the cryptography policy debates, the technological environment is now vastly different than it was when policies were first devised. Leaving aside the advances in telephone switches, we may now also communicate using a number of infrastructures including mobile telephony, Internet, and wireless LAN infrastructures. If governments insist on applying

traditional powers to these new infrastructures, the new policies must acknowledge that the data being collected now is separate from tradition.

Many policy initiatives have involved articulations regarding the importance of being technology-neutral. When the Clinton Administration first announced its intention to update lawful access powers to include cable-based internet connections, they proposed ”ammendments [that] will update the statutes in outmoded language that are hardware specific so that they are technologically neutral” [8]. Meanwhile in the United Kingdom, it was noted in the tempestuous debates in the House of Lords, regarding the Regulation of Investigatory Powers Act (RIP) 2000 that:

> The Earl of Northesk: ”One of the many difficulties I have with the Bill is that, in its strident efforts to be technology neutral, it often conveys the impression that either it is ignorant of the way in which current technology operates, or pretends that there is no technology at all.” [10]

Technology-neutral policy is seen as a way to deal with concerns of governments mandating a specific type of technology. While this is favorable in the case of some policies that affect market developments, technology-neutral lawful access policies may contain hazardous side-effects.

Another reason for technology-neutrality is to ensure that new laws do not need to be passed every time a new technology is invented. However, technology-neutral language may be used to ignore, willful or not, the challenges, risks, and costs to applying powers to different infrastructures.

DEFINING ’TRAFFIC DATA’

International governmental organizations have been working for a number of years to ensure lawful access to traffic data, including the Group of 8, Council of Europe. They have been led, through policy-modeling of or pressure from selected countries, including the United Kingdom and the United States.

The Group of 8, the ‘informal’ economic and foreign policy committee of western governments, formed a senior ‘experts’ group in 1995[1] to develop an international co-operation regime to address transnational organized crime. The Lyon Group has since been active on high-technology surveillance-related policies, including three meetings with industry representatives throughout 2000 and 2001. Arising from that work, the G8 working-definition of traffic data is ”non-content information recorded by network equipment concerning a specific communication or set of communications.” [11]

Meanwhile, the Council of Europe (CoE), the 43-member inter-governmental organization, convened closed meetings since 1997 to develop a multilateral treaty establishing lawful access powers across borders. The CoE Convention on Cybercrime defines traffic data as ”any computer data relating to a communication by means of a computer system, generated by a computer system that formed a part in the chain of communication, indicating the communication’s origin, destination, route, time, date, size, duration, or type of underlying service” [2]. In the convention’s Explanatory Report [3], the CoE states that traffic data should be defined so as ’to not refer’ to the content of a communication; but this is a non-binding interpretation. The CoE and G8 definitions have been criticized as ambiguous and problematic, but no change has been achieved due to the closed processes of these fora.

One is left to wonder what is included and what is excluded by these vague definitions. While subject lines in emails may be content (as they ’refer to’ content [3]), uncertainty arises as to whether the name of files requested (e.g. HTTP requests), URLs (e.g. `http://www.computer.tld`), search parameters, TCP headers, and other such data are considered content or traffic data. A report of a transaction by an individual with server 158.143.95.65 may be considered traffic data; but the name of the web site(s) run on that server may disclose more information (e.g. `aidshelpline.com`). Search parameters in the URLs and the name of files accessed may refer to the content of the communications. If we consider the next generation internet, mobility bindings or routing information included in the IPv6 extended header will include location information. The location information is part of the mobility ’signaling’ protocol and hence fits into the above definitions of traffic data.

Some states have tried to deal with this challenge in their legislative language. The UK’s Regulation of Investigatory Powers Act 2001 went through many iterations, particularl y in the so-called ’Big Browser’ debate, before settling on its final terminology. Traffi c data is defined as data about the source and destination

[1] The ‘G7’ started the Lyon Group as Russia had not yet joined the group, later making it the ‘G8’.

of a transaction, and data abo ut the routing and the tying of separate packets together. This definition is complement ed by the definition of 'communications data': data attached to a transaction provided th at it is used by the network; or exists within logs; or other data that is collected by s ervice providers. However the definitions are also quite clear about the extent of inform ation that qualifies: traffic data does not include URLs per se, and may only include the name of the computers running a service, while the specific resource used qualifies as c ontent, and accorded greater protection. Therefore, the IP address is traffic data, while `http://www.url.tld/file.html` is tantamount to content.

Other states have failed to respect this level of technological awareness. Previous U.S. policy differentiated between traffic data from cable and telephone communications. The Cable Act once protected traffic data to a greater degree than telephone traffic data, as viewing habits were considered sensitive. Now that cable infrastructure is also used for internet communications (which were previously used over telephone lines, and thus traditional laws applied), successive White House administrations worked to erase this cable traffic distinction, finally succeeding with the post-September 11 USA-PATRIOT Act. Rather than deal with the specifics of digital communications media and services, the changes in U.S. law reduces the protections of traffic data for cable internet communications to what had previously existed for telephone communications data. The terminology within the U.S. Code is now ambiguous, lacking supporting documentation with elaborate definitions, and is therefore quite similar to the terms within CoE and the G8 documentation.

This can be interpreted as a boon to law enforcement. According to Attorney General Ashcroft:

> Agents will be directed to take advantage of new, technologically neutral standards for intelligence gathering. (...) Investigators will be directed to pursue aggressively terrorists on the internet. New authority in the legislation permits the use of devices that capture senders and receivers addresses associated with communications on the internet [1].

Traffic data blurs with the content of communications as new communications infrastructures are encompassed under existing practices. The legal protection of this data is reduced as distinctions applied are based on categorical decisions established under older technologies. The separation of content and traffic remains elusive, even in policy language.

Categorical Determinants: The technological

Traffic data under the plain old telephone system was considered derivative, and while informative, it did not necessarily disclose the sensitive details of an individual's life. While the Cable Act protections accepted that the data discloses the viewing preferences of individuals and therefore deserved greater protections, such protections were later deemed irrelevant for the internet.

Traffic data's constitution differs by communications medium. Below we present dial-in records, wireless LANs, and search engines to preview what can accessed by technology-neutral law enforcement powers.[2]

Dial-In records

The Remote Authentication Dial-In User Service (RADIUS) is a client/server security protocol, designed to manage dispersed modem pools for large numbers of users. This tends to involve managing a single "database" of users, which allows for authentication (verifying user name and password) as well as configuration information detailing the type of service to deliver.

Many Internet Service Providers are outsourcing the access network to big operators that provide dial-up connectivity world-wide. Internet users dial into a modem pool attached to a Network Access Server (NAS) that operates as a client of RADIUS. The client is responsible for passing user information to designated RADIUS servers (managed by the ISP) and then acting on the response that is returned.

The RADIUS server stores usage information for dial-in users, often for billing purposes. When the user is authenticated and the session has been configured according to the authorization information, an accounting start record is created. When the user's

[2]The data presented has been obtained with permission from a telephone carrier, an internet service provider, and a large conference where wireless LAN access was provided. All transactions presented in this paper have been de-identified, and the time-logs were altered to reduce the risk of re-identification.

session is terminated, an accounting stop record is created.

The most significant fields of the ”start/stop” records are:

- *Timestamp:* Timestamp records the time of arrival on the RADIUS accounting host measured in seconds since the epoch (00:00 January 1, 1970 GMT). To find the actual time of the event, subtract Acct-Delay-Time from Timestamp.

- *Call(ed,ing)-Station-Id:* Where the Called-Station-Id records the telephone number called by the user and the Calling-Station-Id records the number the user is calling from. This information is recorded when the NAS-Port-Type is ISDN, ISDN-V120, or ISDN-V110.

Start and stop RADIUS records may look like:

[See Appendix I.B]

From this log we can extract a limited amount of information regarding the content of the communications transactions that took place. The user has been identified (aep@somedomain.org), the number of the caller (01223555111, which is a Cambridge number) and the place being called (02075551000, London), IP address assigned (62.188.17.227), the duration (21s), type of connection, date and time. The traffic data over time identifies the change in location of a user despite the common dialed number. As users roam globally with different access telephone numbers, the user identification remains static. In this sense, the collected traffic data is mildly more sensitive than traditional telephone data: where POTS traffic data pivots around a given telephone/ID number, RADIUS data pivots around a user ID regardless of location; therefore disclosing location shifts.

Wireless LAN association records

Such mobility becomes more problematic within wireless environments. In a standard wireless LAN environment using IEEE 802.11b, a radio cell size can vary from hundreds of meters in open air, to a small airport lounge. Before the mobile station (STA) is allowed to send a data message via an access point (AP), it must first become associated with the AP. The STA learns what APs are present and then sends a request to establish an association.

The significant records of a centralized association system log are:
- *time_GMT:* Time when a mobile node associates with a base station
- *Cell_ID:* Base station unique identifier in the LAN
- *MAC_ID:* Media Access Control address identifier; a unique Identifier of a mobile device

[See Appendix I.C]

It is tempting to analyze these logs by drawing an analogy with the POTS, i.e. a registration of a mobile with an access point could be seen as the establishment of a phone call between both parties. This analogy is simplistic as it doesn’t consider that the Cell_IDs represent places (airport, conference room, restaurants) and the registration timestamps can reveal if two nodes are (moving) together. Data mining of association records (registration and deregistration) can provide sufficient information to draw a map of human relationships. [4]

HTTP requests to a search engine

The above media may involve further traffic data in the form of internet protocols. The GET and POST methods in the Hypertext Transfer Protocol (HTTP) allow a web client to interact with a remote server. In the most common search engines, the keywords are included in the HTTP header as part of a GET method. All the web logs can be transformed to a W3C common log file format that contains the IP address of the client, the connection time, the object requested and its size.

[See Appendix I.D]

If ’traffic data’ residing in logs are accessed by authorities, a great deal of intelligence can be derived. Observing the logs we can see for example, that 212.164.33.3 has requested (in a short period of time) information about ”railway+info+London” and ”union+strike” in two different requests. This is the ability to find out not only the patterns of an individual’s movements on-line, but also to identify an individual’s intentions and plans. Or more dangerously one could derive false intentions (child+pornography may be a search for studies on the effects of pornography on children). Much more can be ascertained with some datamining, even if IP addresses are assigned dynamically, allowing for traceability based on habits

and interests; and compounded with location data, previous NAS data, etc., a comprehensive profile can be developed.

THE SHAPE OF THINGS…

Even the Council of Europe acknowledges, within the convention's Explanatory Report [3], that the breadth of possible traffic data may be problematic.

> "The collection of this data may, in some situations, permit the compilation of a profile of a person's interests, associates and social context. Accordingly Parties should bear such considerations in mind when establishing the appropriate safeguards and legal prerequisites for undertaking such measures."

No such safeguards or prerequisites are discussed in detail, or mandated as, again, the Explanatory Report is non-binding and often ignored.

Shifting between infrastructures gives different data; but converging infrastructures is even more worrisome. Mobile communications systems magnify the sensitivity of traffic data; wireless LANs were presented as an indication of the shape of things to come as we encounter new protocols and infrastructures, e.g. third generation wireless running IPv6.

This exposition of traffic data could be extended to mobile telephony; and to understanding the output of devices such as Carnivore (DCS1000); the point can be made that the data collected depends on both the infrastructure and the means of collection. The collection and access methods currently under consideration are preservation (access to specified data of a specific user that are collected by service providers for business purposes), retention (requiring all logs for all users be stored beyond their business purpose for government access), and real-time (governmental access to real-time data flows).

The national laws that enshrine these access powers differ remarkably, despite being established under the umbrella/guidance of international organizations such as the G8 and the CoE. The UK appears to separate URLs from traffic data; but in the same piece of legislation assured that ministers sign interception warrants, and in later policies and legislation proposed retention regimes for periods of time ranging from 4 days (web cache), 6 months (RADIUS, SMTP, and IP logs), and 7 years [5]. The U.S. recently introduced technological-neutrality to its laws thus reducing protections; but the U.S. does require judicial warrants for interception, and has no retention requirements. The CoE convention places no requirements on countries to require judicial authorizations, and with 33 signatory states including the U.S., Canada, South Africa, Romania, France, and Croatia, we can rest assured that there will be selective interpretation in implementation. Even among the G8 countries, the protections afforded to citizens' communications in Italy, Germany, the U.S. and Russia vary greatly. Already the Canadian government has proposed, in its efforts to ratify the CoE convention, to consider all telecommunications services as equivalent, and argues that "the standard for Internet traffic data should be more in line with that required for telephone records and dial number recorders in light of the lower expectation of privacy in a telephone number or Internet address, as opposed to the content of a communication." [6]

While policies may vary within and across borders, the nature of the data produced and its sensitivity does not. 'Traffic data' analysis generates more and more sensitive profiles of an individual's actions and intentions, arguably more so than communications content. In a communication with another individual, we say what we choose to share; in a transaction with another device, e.g. search engines and cell stations, we are disclosing our intents, actions, and movements. Policies continue to regard this transactional data as plain old telephone system 'traffic data', and accordingly apply old protections.

This is not faithful to the spirit of updating laws for new technology. We need to acknowledge that changing technological environments alter the habitat of a policy. New policies need to reflect the totality of the new environment.

The technology policy innovations fail to do so. Governments seek technology-neutral policy, and are also doing so at the international level. This appears to be to the advantage of policy-setters. New powers are granted through technological ambiguity rather than clear debate and due process. International instruments, such as those from the Group of 8 and the Council of Europe, harmonize language in a closed way with little input and debate. This problem will grow as more countries feel compelled to ratify and adopt these instruments; or feel that it is in their interests to do so.

Attempts to innovate policy must be interrogated, lest we reduce democratic protections and oversight blindly.

About the Authors

Ian (Gus) Hosein is a Visiting Fellow in the Department of Information Systems at the London School of Economics; and a Senior Fellow at Privacy International. For more information please see http://is.lse.ac.uk/staff/hosein/

Alberto Escudero Pascual is a Research Assistant in the Telesystems Laboratory at the Royal Institute of Technology (KTH) in the area of privacy in the next generation Internet. For more information please see http://www.it.kth.se/~aep/

References

[1] Ashcroft, J. Testimony of the Attorney General to the Senate Committee on the Judiciary. Washington D.C. September 25, 2001.

[2] Council of Europe. Convention on Cybercrime, ETS no.185, opened for signature on November 8, 2001. http://conventions.coe.int/

[3] Council of Europe. Convention on Cybercrime Explanatory Report, adopted on November 8, 2001. http://conventions.coe.int/

[4] Escudero A. Contribution to the EU Forum on cybercrime. Location data and traffic data. Brussels. November 2001.

[5] Gaspar, R. Looking to the Future: Clarity on Communications Data Retention Law: A National Criminal Intelligence Service submission to the Home Office for Legislation on Data Retention. Submitted on behalf of ACPO and ACPO(S); HM Customs & Excise; Security Service; Secret Intelligence Service; and GCHQ, August 2000.

[6] Government of Canada. Lawful Access - Consultation Document. Department of Justice, Industry Canada, Solicitor General Canada. August 25, 2002.

[7] Hosein, I., and Whitley, E. ”Developing national strategies for electronic commerce: Learning from the UK’s RIP Act.” Journal of Strategic Information Systems, Volume 11, Number 1, 2002.

[8] Podesta, J. National Press Club Speech with (former) White House Chief of Staff John Podesta on ”Cyber Security”. Washington D.C. July 17, 2000.

[9] Reno, J. Law Enforcement in Cyberspace Address by The Honorable Janet Reno, (former) United States Attorney General. San Francisco: Presented to the Commonwealth Club of California, 1996

[10] UK Hansard. ”House of Lords 28th June, 2000 (Committee Stage)”, Column 1012 (published by The Stationery Office Limited).

[11] U.S. Delegation to G8. Discussion Paper for Data Preservation Workshop. Tokyo, G8 Conference on High-Tech Crime. May 22-24 2001.

I. Appendix

A. CDR

The call data records look like:

```
19991003070824178 165 0187611205 46732112106 ----------001------003sth 46 4673000-----0013 14 10260
1999100307083041 33 01541011341 46708314801 ----------001------003sth 46 4670000--8 0013 11 10260
1999100307162963 51 0187614815 46739112106 ----------001------003sth 46 4673000-----0013 13 10260
1999100307182788 74 015410124301 46708314801 ----------001------003sth 46 4670000--8 0014 11 10260
1999100307204736 18 0187614805 46739112106 ----------001------003sth 46 4673000-----0013 14 10260
1999100307222326 20 01317023888 46706263087 ----------001------003sth 46 4670000--6 0013 1 10260
1999100300131791 90 0131654200 46854543084 ----------001------002sth 46 46 001-----0014 14 10260
```

Fig. 1. Call Data Records

B. Radius Records

A start and stop radius records looks like:

▹ ***Fri Oct 19 11:30:40 2001***
User-Name = "aep@somedomain.org"
NAS-IP-Address = 62.188.74.4
NAS-Port = 3239
NAS-Port-Type = Async
Acct-Status-Type = Start
Acct-Delay-Time = 0
Acct-Session-Id = "324546354"
Acct-Authentic = RADIUS
Calling-Station-Id = "01223555111"
Called-Station-Id = "02075551000"
Framed-Protocol = PPP
Framed-IP-Address = 62.188.17.227
Proxy-State =
"PX01\0\0'\0xcdntg\0x13\0xfe\0xfe\0xdd+ew\0xdfV\0xa4\0xc7Y\[...]\0xfc\0x8c"

▹ ***Fri Oct 19 11:31:00 2001***
User-Name = "aep@somedomain.org"
NAS-IP-Address = 62.188.74.4
NAS-Port = 3239
NAS-Port-Type = Async
Acct-Status-Type = Stop
Acct-Delay-Time = 0
Acct-Session-Id = "324546354"
Acct-Authentic = RADIUS
Acct-Session-Time = 21
Acct-Input-Octets = 11567
Acct-Output-Octets = 3115
Acct-Input-Packets = 96
Acct-Output-Packets = 74
Calling-Station-Id = "01223461172"
Called-Station-Id = "9061000"
Framed-Protocol = PPP
Framed-IP-Address = 62.188.17.227
Proxy-State = "PX01\0\0\0x1b\0x93;\0xaa\0x98\0xea\0xad\[...]\0xc7\0xffi"

Fig. 2. Radius Data Records

From the previous log we can extract the following information:
User: aep@somedomain.org
Place of call: Cambridge (UK) 01223555111
Calling to: London (UK) 02075551000
IP address: 62.188.17.227
Durantion of call: 21 Seconds
Type of connection: ASYNC MODEM
Date and time: from Fri Oct 19 11:30:40 2001 to Fri Oct 19 11:31:00 2001

C. WLAN Authentication Records

WLAN authentication records looks like:

```
time_GMT=20010810010852 Cell_ID=115 MAC_ID=00:02:2D:20:47:24
time_GMT=20010810010852 Cell_ID=115 MAC_ID=00:02:2D:04:29:30
time_GMT=20010810010852 Cell_ID=115 MAC_ID=00:60:1D:21:C3:9C
time_GMT=20010810010853 Cell_ID=129 MAC_ID=00:02:2D:02:40:EF
time_GMT=20010810010854 Cell_ID=129 MAC_ID=00:02:2D:1F:53:C0
time_GMT=20010810010854 Cell_ID=129 MAC_ID=00:02:2D:09:17:E8
time_GMT=20010810010854 Cell_ID=129 MAC_ID=00:02:2D:1D:67:FE
time_GMT=20010810010856 Cell_ID=41 MAC_ID=00:02:2D:0A:5C:D0
time_GMT=20010810010856 Cell_ID=41 MAC_ID=00:02:2D:1F:78:00
time_GMT=20010810010856 Cell_ID=41 MAC_ID=00:60:1D:1E:D4:53
time_GMT=20010810010858 Cell_ID=211 MAC_ID=00:60:1D:F0:E4:D8
time_GMT=20010810010900 Cell_ID=154 MAC_ID=00:30:65:00:62:27
time_GMT=20010810010900 Cell_ID=154 MAC_ID=00:02:2D:05:0B:25
time_GMT=20010810010900 Cell_ID=154 MAC_ID=00:60:1D:22:26:A7
time_GMT=20010810010900 Cell_ID=154 MAC_ID=00:02:DD:30:06:90
time_GMT=20010810010900 Cell_ID=154 MAC_ID=00:02:2D:0D:27:D3
```

Fig. 3. WLAN Authentication Records

D. HTTP Search Engine Queries

Extracting search queries from a web log we can obtain records that look like:

```
295.47.63.8 - - [05/Mar/2002:15:19:34 +0000] "GET /cgi-bin/htsearch?confi g=htdigx&words=startrek HTTP/1.0" 200 2225
295.47.63.8 - - [05/Mar/2002:15:19:44 +0000] "GET /cgi-bin/htsearch?confi g=htdig&words=startrek+avi HTTP/1.0" 200 2225
215.59.193.32 - - [05/Mar/2002:15:20:17 +0000] "GET /cgi-bin/htsearch?confi g=htdig&words=Modem+HOWTO HTTP/1.1" 200 2045
192.77.63.8 - - [05/Mar/2002:15:20:35 +0000] "GET /cgi-bin/htsearch?confi g=htdig&words=conflict+war HTTP/1.0" 200 2225
211.164.33.3 - - [05/Mar/2002:15:21:32 +0000] "GET /cgi-bin/htsearch?confi g=htdigx&words=railway+info HTTP/1.0" 200 2453
211.164.33.3 - - [05/Mar/2002:15:21:38 +0000] "GET /cgi-bin/htsearch?confi g=htdigx&words=tickets HTTP/1.0" 200 2453
211.164.33.3 - - [05/Mar/2002:15:22:05 +0000] "GET /cgi-bin/htsearch?confi g=htdigx&words=railway+info+London HTTP/1.0" 200 8341
212.164.33.3 - - [05/Mar/2002:15:22:35 +0000] "GET /cgi-bin/htsearch?confi g=htdigx&words=union+strike HTTP/1.0" 200 2009
82.24.237.98 - - [05/Mar/2002:15:25:29 +0000] "GET /cgi-bin/htsearch?confi g=htdigx&words=blind+date HTTP/1.0" 200 2024
```

Fig. 4. HTTP Search Engine Records

Observing the logs we can see for example, that *212.164.33.3* has requested (in a short period of time) information about “railway+info+London” and “union+strike” in two different requests.

PAPER #7

Alberto Escudero-Pascual

"Privacy in mobile internet in the context of the European Union data protection policy"

Internet Society Conference (INET2002).
Washington DC, USA
June 2002

EUROPEAN UNION DATA PROTECTION POLICY 'LOCATION PRIVACY IN THE NEXT GENERATION MOBILE INTERNET'

Alberto Escudero Pascual <aep@kth.se>
Royal Institute of Technology. KTH
IMIT - IT University
Kista. Stockholm. Sweden

Abstract - The global telecommunication infrastructure will slowly converge toward an integrated packet switched network using the Internet Protocol as the common communication technology. First evidence of this convergence is the deployment of the third generation wireless infrastructure that brings together the radio access network and core network by using the next generation Internet Protocol IPv6.

The paper presents how 'mobility' is supported in IPv6. Mobility is the capability of mobile terminal to be reachable by its home network IP address with independence of the point of attachment to the Internet. We show the kind of information items that are required to be in transit in the network to allow a mobile node to seamlessly communicate on the move.

The European Commission proposal for a Directive (COM(2000)385) on 'processing of personal data and protection of privacy in the electronic communication sector' is kept in a technological neutral manner which means that no standards are imposed. It includes definitions inter alia on 'location data' and 'traffic data' and foresees privacy safeguards and different levels of protection for distinct kind of data. After the technical and legal overview we discuss the difficulties to apply the definitions provided by the Directive to certain technology as mobility in IPv6.

Based on the reasoning included in the paper we argue that classifying and defining data by traditional means and ways without taking into account Internet's multi-layered architecture might lead to an insufficient level of privacy protection for certain sensitive data and might not be the most appropriate way to adapt and update the existing provisions to new and foreseeable developments in electronic communications services and technologies.

INTRODUCTION

This paper is divided as follows: Section 1 contains a brief overview of the Internet Protocol Version 6 and how mobility is supported, Section 2 introduces the proposed European Union Directive COM(2000)385 concerning the processing of personal data and the protection of privacy in the electronic communication sector with special remark to the privacy protection of location and traffic data, Section 3 presents some of open issues when trying to apply the Directive in the context of IPv6 mobility concerning the interpretation of the 'technology-neutral' definitions of traffic, content and location data. Finally, Section 4 presents some conclusions and regulatory recommendations.

I. THE NEXT GENERATION INTERNET

The Internet Protocol Version 6 (IPv6), also known as "IPng" (IP Next Generation), is the latest version of the Internet Protocol (IP). Formally, IPv6 is a set of specifications from the Internet Engineering Task Force (IETF). IPv6 is being designed as an evolutionary set of improvements to the current IP Version 4. The most obvious improvement in IPv6 over the IPv4 are that IP addresses are lengthened from 32 bits to 128 bits which anticipates the future growth of the Internet and provides relief for what was perceived as an impending shortage of network addresses. Besides, IPv6 offers technical advantages over IPv4, including self-configuration mechanisms, enhanced security, quality of service features and native mobility support [1]. IPv6 aims to be the protocol capable of bringing together access and core networks, the 'glue' for the deployment of the future 'all-IP' telecommunication network.

IPv6 includes a security protocol in the network layer that provides cryptographic security services that supports combinations of authentication, integrity, access control, and confidentiality. The IP

Encapsulating Security Payload (ESP) and the IP Authentication Header (AH) are part of the IP Security architecture (IPSEC) described in RFC 2401[2].

Both ESP and AH are mandatory parts of IPv6 and make sure that a third party eavesdropping on the channel can not read and/or modify the IP datagram. The IP Authentication Header seeks to provide security by adding authentication information to each IP datagram whilst confidentiality requires the use of ESP. Neither AH nor ESP hide the source and destination IP addresses of the communicating parties and hence their network location.

A. Mobility support in IPv6

The protocol operation defined for mobility in IPv6 is known as MobileIPv6 [3] and allows a mobile node to move from one link to another without changing the mobile node's IP address. A mobile node is always addressable by its 'home address' (HoA), an IP address assigned to the mobile node within its home subnet, i.e., with the prefix of its home link.

MobileIP allows users to move between different networks, while maintaining an addressable static identifier (home address). This is done by associating a dynamic identifier (care-of-address) with the mobile node when it is away from home. All traffic to the mobile node is intercepted in the home network by a home agent (HA) that tunnels the data to the care-of-address that is in use in that moment. Packets may be routed to the mobile node using their home address regardless of the mobile node's current point of attachment to the Internet (CoA), and the mobile node may continue to communicate with other nodes after moving to a new link. With MobileIP the movement of a mobile node away from its home link is thus transparent to transport and higher-layer protocols and applications.

MobileIPv6 shares many features with MobileIPv4[4], but the protocol is now fully integrated into IPv6. MobileIPv6 works in a similar way as MobileIPv4 does when using mobile node co-located care of address mode and route optimization. As in MobileIPv4 the mobile mode is responsible for discovering its current location. When the mobile mode is attached to its home link it directly receive packets and when roaming in a foreign network, it must acquire a co-located care of address and notify its home agent of this address.

MobileIPv6 on the other hand also includes the Mobility Header a new IPv6 protocol that allows the mobile node to inform selected IPv6 correspondent hosts of its care-of-address, so packets from these correspondent hosts can be redirected straight to the mobile node instead of using the home agent as an intermediary. The association between a mobile node's home address and care-of address is known as a 'binding' for the mobile node. A mobile node typically acquires its care-of address through stateless or stateful (e.g., DHCPv6) address autoconfiguration and while away from home registers its care-of address with a router on its home link, requesting this router to function as his home agent.

1) Binding dynamic (CoA) and static identifiers (HoA): This binding registration is done by the mobile node sending to the home agent a packet with a Mobility Header containing a Binding Update message; the home agent then replies to the mobile node by returning a packet containing a Binding Acknowledgment message.

The Mobility Header by a set of different Binding (Request, Acknowledgment, Update and Missing) messages allows correspondent hosts communicating with a mobile node, to dynamically learn and cache the mobile node's binding (HoA-CoA).

Before sending a packet to any IPv6 destination, a node checks its cached bindings for an entry for the packet's destination address. If a cached binding for this destination address is found, the node uses an IPv6 Routing header to route the packet to the mobile node by way of the care-of address indicated in the cached binding. No IPv6 encapsulation is required due to IPv6's routing header. If a mobile node is at home destination the mobile node sees the home address in the routing header and processes the packet. If the sending node has no cached binding for the destination address, the node sends the packet normally to the home address without any Routing header and the packet is subsequently intercepted and tunneled by the mobile node's home agent. The correspondent nodes can also send a Binding Request message to the mobile node home address in order to obtain its care-of-address.

In summary mobility is enabled in IPv6 by two basic functions:

- **Mobility Header:** A mechanism to keep informed to the home agent and correspondent nodes of the changes of the mobile node's point-of-attachment to the Internet.

- **Efficient Routing:** Enable by including the home address in the destination option of *each* of the packets send by the mobile node to the correspondent node and

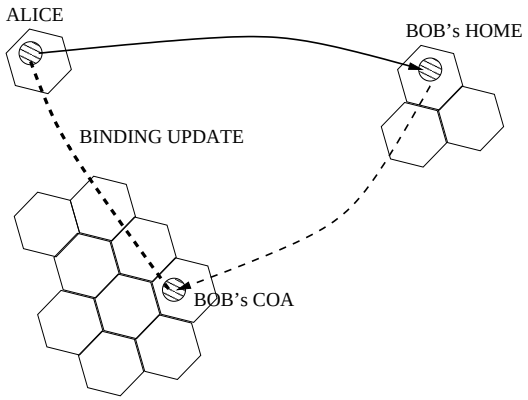

Fig. 1. MobileIPv6. Node reachability

by including the mobile node's home address in the routing header when the correspondent node talk to the mobile node via its care-of-address.

In the first case the binding process makes use of certain protocol named Mobility Header while the efficient routing is achieved by including the home address of the mobile node *as part of each* of the packets exchanged.

II. EU Data Protection Directive

On July 12, 2000 the European Commission adopted a proposal for a Directive (COM(2000)385) on 'processing of personal data and protection of privacy in the electronic communication sector'. The proposal is part of a package of proposals for initiatives which will form the future regulatory framework for electronic communications networks and services. It aims to adapt and update the existing Data Protection Telecommunications Directive (97/66/EC) to take account of technological developments.

Unlike the previous Telecommunications Directive the scope of the future Directive would not be restricted to telephony and data networks, but will also cover satellite, terrestrial and cable TV broadcasting networks, irrespective of the type of information concerned.

The new proposed EU Data Protection Directive establishes a common framework for data protection in telecommunication services and networks regardless of the technology in use in electronic communication services and networks. The Directive provides a set of protections or safeguards and various definitions such as: communication channel, traffic and location data, service provider, etc.

As part of the proposed changes, the new Directive replaces the existing definitions of telecommunications services and networks in Directive (97/66/EC) to align the terminology with a common framework for electronic communications services and networks. The changes of definitions try to ensure that all kind of electronic communications will be covered regardless of the technology in used. The Directive also include four new important definitions to strengthen the common understanding. They relate to the following:

- **Communication** is any information exchanged or transmitted between a finite number of parties by means of a publicly available electronic communications service.
- **Call** is a connection established by means of a publicly available telephone service allowing two-way communication in real time.
- **Traffic Data** is any data generated and processed in the course of or for the purpose of the transmission of a communication over an electronic communications network.
- **Location Data** is any data processed in an electronic communications network, indicating the geographic position of the terminal equipment of a user of a publicly available electronic communications service.

For the purpose of this paper we consider the protections related to traffic and location data. The Article 6 of the directive prohibits the use of traffic data expect for billing purposes and introduces the possibility for further data processing for value-added services based on consent of user/subscriber. Article 9 introduces specific privacy safeguards for subscribers and users with regard to mobile location information services.

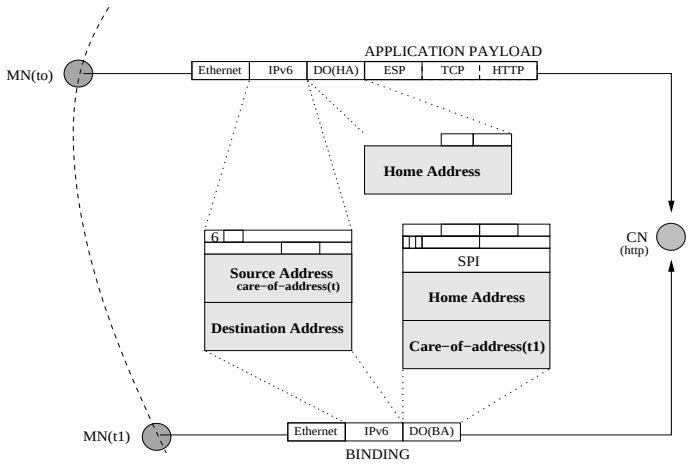

Fig. 2. Mobility/Location Information embedded in IPv6 headers

III. Open issues in mobile privacy

In this section we present two issues that arises when trying to apply the Directive to the concrete scenario of mobility in IPv6. The first one concerns the distinction between traffic and content of the communication in the Internet and the second is related to the meaning of geographical position.

A. From traffic data to content data

Traditionally content data is considered to be more sensitive than traffic data and therefore embedded in a higher level of privacy protection. But on the Internet content and traffic data can only be clearly distinguished when given a very concrete context, item of interest and level of observation [8]. Internet Protocols take benefit of a multi-layered architecture, where higher-level components talks to lower-level components and vice-versa. The result of this layered architecture is that what can be seen as traffic/signaling for one layer can be indeed be content for a lower layer.

The benefits of a layered architecture is that the communications functions are partitioned into a set of hierarchical layers where each layer performs a related subset of the functions required to communicate relying on the next lower layer to perform more primitive functions while provides services to the next higher layer.

But it is not only the Internet's multi-layered architecture which makes a clear distinction between content and traffic data difficult. Moreover, the issue of interest plays a fundamental role in the discussion, what can be pure traffic or signaling data for one observer can be considered content for another. For example, the fact that two parties are communicating in a given time can provide an observer with a certain amount of sensitive information not requiring to examine the 'formal content' of the communication at all.

Another clear example of this complexity can be found in mobility in IPv6. As shown in [I-A] MobileIPv6 supports transparency above the IP layer that includes the maintenance of active TCP connections when a mobile node changes its point of attachment to the Internet. From the point of view of the application, the mobile node always use the mobile node's home address as being statically in the home network.

The binding messages can be seen as pure 'mobility signaling' hidden to the application or on the contrary as a rich content information that allows to the mobile node to make location aware decisions.

In its provisions the Directive tries to adapt the traditional mechanism to differentiate signaling and content in traditional plain old telephone systems (POTS) to the Internet or other packet switched networks and services. In POTS signals and data are transported in different channels and include a very restricted number of message formats. The signaling channels take care of the interactions between the components. Typically, this information is either data being transferred between end-users and other end-users (known as content) or between end-users and the network (traffic).

The traditional way of classifying the data conflicts with the pure nature of a multi-layered architecture of the Internet where the distinction depends on the functionality and layer of observation. What can be

seen as one communication process from the point of view of the user (exchange of e-mail between two mail agents) involves multiple levels of signaling and content.

B. From traffic data to geographical position

The second open question is related to definition of geographical position and the Internet addressing.

As shown in [I-A.1] the mobility header and the efficient routing make sure that the communicating parties are aware of the changes of the mobile node attachment to the Internet. Opposite to the old communication systems, where signaling and content are transmitted in certain channels, what could be considered as 'IPv6 mobility signaling' is a set of IP packets or/and packet's fields. [Fig.2] shows the structure of two 'mobileipv6' packets exchanged between a mobile node and a web server, the first one is sent during the web browsing session and the second is a 'binding update'. The figure also shows how mobility related information is present in both packets and embedded in two different ways in the IPv6 header.

When an eavesdropper is located somewhere along the route between the home agent and the mobile node it is possible to identify and track the mobile node movements by observing the 'mobility signaling' exchanged between the different communicating parties. The mobility of a certain mobile node [Fig.1] with home address HoA_i in a period of time $(t_o < t < t_n)$ can be represented as a series of care-of-addresses $[CoA(t_o), CoA(t_1), CoA(t_2)...CoA(t_n)]_i$.

It is out of the scope of this paper to describe the different Geographic Information Reference Systems but while the care-of-address doesn't represent a standardized geographical position due to the nature of the reference system (the Internet infrastructure), the care-of-address and its changes during the time can without any doubt reveal mobility and if not absolute, relative positioning [7].

Part of the ongoing research work of the *PETng Project [6]* is to create a model that will allow to an eavesdropper to determine the proximity of two mobile nodes by observing the changes of their CoA(t) during the time i.e. to determine the 'relative positioning' of two mobile nodes.

IV. Conclusions

Traditional legal, regulatory and technical provisions were established with traditional technological environments in mind. When telephone traffic data was decided to be less invasive than the content of the conversation, this reflected the plain-old-telephone system (POTS): traffic data was merely the person who was calling or the person called, and the duration. Accordingly, one level of privacy protection was typically assigned to traffic data (if at all), and another was applied to access to communications content, that is, the conversation itself.

The traditional classification of data based on which functional channel is sent and received conflicts with the pure nature of a multi-layered architecture of the Internet where the distinction depends on the layer of observation and the interactions that are observed.

In order to illustrate the complexity of using technology-neutral language we have shown how problematic is to classify the data in traffic, content and location in mobile Internet. What it can be considered as traffic data can be either content or location depending on the items of interest for the observer and/or application.

It is our concern that technology-neutral language may be used to ignore, willful or not, the challenges, risks, and costs to applying powers to different technical infrastructures. Classifying and defining data by traditional means and ways without taking into account Internet's multi-layered architecture might lead to an insufficient level of privacy protection for certain sensitive data based on the fact that they are grouped under certain category.

If policy makers insist on applying traditional powers to these new infrastructures, we argue that the new lawful policies must acknowledge that the data being collected now is separate from tradition and can only be understood having the different technologies in mind.

V. Acknowledgments

I will like to acknowledge to Ian (Gus) Hosein from the Department of Information Systems at the London School of Economics with whom i started to research in this interdisciplinary area and Corinna Schulze from the Directorate General Information Society of the European Commission for giving background information on the existing legal Framework on Data Protection.

Lastly i want to express my gratitude to the Personal Computing and Communication (PCC) research program and the Swedish Center for Internetworking (SCINT) for supporting my work in Internet privacy.

REFERENCES

[1] **C. Huitema,** *IPv6, the new Internet Protocol.* 2nd Edition. Prentice Hall. 1997.

[2] **S. Kent and R. Atkinson**, *"Security Architecture for the Internet Protocol"*. RFC 2041.

[3] **D. Johnson and C. Perkins,** *"Mobility Support in IPv6"*, draft-ietf-mobileip-ipv6 v16. March 2002.

[4] **C. Perkins,** *"IP Mobility Support"*, RFC 2002, October 1996.

[5] **A. Escudero**, *"Anonymous and Untraceable Communications: Location Privacy in Mobile Internetworking"*. Licentiate Thesis. July 2001.

[6] **PETng Project.** *"Privacy Enhanced Technologies in the next generation Internet"*. SCINT/KTH/KAU. http://www.petng.net

[7] **A. Escudero**, Contribution to the EU Forum on cybercrime. Location data and traffic data. Brussels. November 2001.

[8] **I. Hosein, A. Escudero,** *"Understanding traffic data and deconstruction technology-neutral regulations"*. UNECE. March 2002.

PAPER #8

Alberto Escudero-Pascual, Thijs Holleboom and Simone Fischer-Huebner

"Privacy for location data in Mobile Networks"

Nordic Security Workshop (NORDSEC2002), pp. 220-232
Karlstad, Sweden
November 2002

PRIVACY FOR LOCATION DATA IN MOBILE NETWORKS

A. Escudero-Pascual
<aep@kth.se>
IMIT
IT University - KTH
Stockholm. Sweden

T. Holleboom
<thijs@cs. kau.se>
Computer Science Dept.
Karlstad University
Karlstad. Sweden

S.Fischer-Huebner
<simone@cs.kau.se>
Computer Science Dept.
Karlstad Universit y
Karlstad. Sweden

Abstract - The new EU Directive 2002/58/EC has introduced with its Art. 9 special protection for location data other than traffic data. In this paper, we argue that also location data within traffic data can contain sensitive information about the ”relative positioning” and ”co-located displacements” of mobile nodes and thus also requires special protection.

After a brief introduction to how mobility is supported in IP networks, to the level of privacy protection for location data introduced in the new European Union data protection directive, and to means of protecting privacy by technology, we introduce the concept of *co-located displacements in MobileIP* and show how the home agent will be able to determine whether or not a set of mobile nodes move in a co-located fashion.

Finally, we present how privacy-enhancing technologies can be used to provide the level of privacy protection as required by Art. 9 of the EU Directive 2002/58/EC for location data other than traffic data, also for location information within traffic data.

INTRODUCTION

Location-based services (LBS) can be described as applications that exploit knowledge about where an information device (user) is located. For example, location information can be used to provide automobile drivers with optimal routes to a geographical destination or inform a group of friends when or where a friend is close in the neighborhood.

Traditionally security in computer networks include different aspects of message integrity, authentication, and confidentiality. However, in wireless networks, where users move between different networks and media types, another issue becomes equally important: location privacy.

This paper focuses on the situation where the absolute or relative position is computed in the infrastructure (i.e., home agent) in MobileIP-based networks.

The actual task of a home agent on a mobile node’s home network is to tunnel datagrams for delivery to the mobile node when it is away from home (see section I.A). In the future, however, it might become more and more common that mobile nodes are also offering value-added services, such as LBS.

In these scenarios the user is not in full control of the location information associated with the mobile device. The problem arises when location information is required in order to obtain a service and at the same time the user does not want to reveal more personal identifiable information than is strictly necessary for the provision of a concrete service. For example, a mobile user may want to inform to only a certain number of people for a certain period of time about his or her position or, to learn the position of the nearest catholic church without revealing his or her personal identity.

The paper is divided as follows:

Section 1 gives an introductory overview to mobility in IP networks, the European Union data protection directive concerning the processing of location data and to privacy-enhancing technologies useful for protecting location data, such as the Platform for Privacy Preferences Protocol (P3P) and *mix nets*.

Section 2 describes co-located displacements in MobileIP and proposes a formal method that allows a home agent to determine whether or not two mobile nodes move in a co-located fashion.

Section 3 explains how mix nets can be used to anonymise location information and how to use P3P to technically support the legal requirement of informed consent for the processing of location data within traffic data for value-added services.

I. BACKGROUND

A. Mobility support in IP networks

The protocol operation defined for mobility in IP networks is known as MobileIP[1]. MobileIP allows a mobile node to move from one link to another in the Internet without changing the mobile node's home IP address. With MobileIP the mobile node can seamlessly roam among IP networks and media types without restarting any of the ongoing connections or associated applications. A mobile node is always addressable by its home address (HoA), an IP address assigned to the mobile node within its home subnet, i.e., with the network prefix of its home link.

MobileIP allows users to move between different networks, while maintaining an addressable static identifier (home address). This is done by associating a dynamic identifier (care-of-address, CoA) with the mobile node when it is away from home at a foreign link. All traffic to the mobile node is intercepted in the home network by a home agent (HA) that tunnels the data to the care-of-address that is in use in that moment. Packets may be routed to the mobile node using their home address regardless of the mobile node's current point of attachment to the Internet (CoA), and the mobile node may continue to communicate with other nodes after moving to a new link. With MobileIP the movement of a mobile node away from its home link is thus transparent to transport and higher-layer protocols and applications.

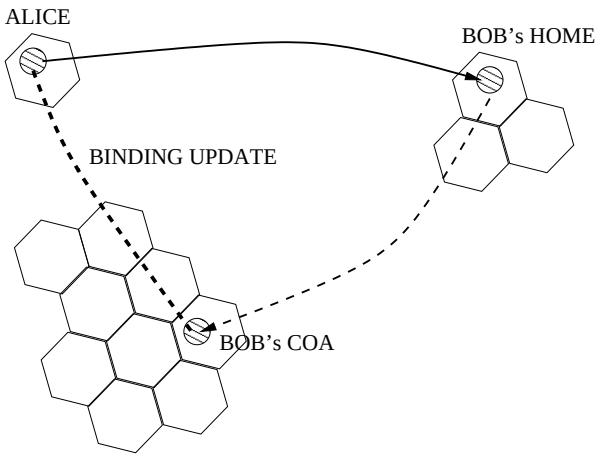

Fig. 1. Route Optimization in MobileIPv6.

MobileIPv6 shares many features with MobileIPv4, but the protocol is now fully integrated into IPv6. As in MobileIPv4 the mobile mode is responsible for discovering its current location. When the mobile mode is attached to its home link it directly receives packets and when roaming in a foreign network, it must acquire a co-located care of address and notify its home agent of this address.

MobileIPv6 on the other hand also includes the mechanisms that allows the mobile node to inform selected IPv6 correspondent nodes (CN) of its care-of-address, so packets from these correspondent hosts can be redirected straight to the mobile node instead of using the home agent as an intermediary (route optimization) [Fig .1].

B. European Union Directive on privacy in electronic communication

On July 12, 2002 the EU Directive 2002/58/EC concerning 'processing of personal data and protection of privacy in the electronic communication sector' [2] was officially adopted. The new Directive is part of a package of initiatives which will form the future regulatory framework for electronic communications networks and services. It aims to adapt and update the existing Data Protection Telecommunications Directive (97/66/EC) [3] to take account of technological developments.

The new EU Directive 2002/58/EC establishes a common framework for data protection in telecommunication services and networks regardless of the technology in use in electronic communication services and networks.

Whereas in the Directive 97/66/EC traffic data only refers to "calls" in so-called circuit switched connections (traditional voice telephony or plain old telephone system aka. POTS), the new EU Directive 2002/58/EC covers all traffic data in a technology neutral way including Internet traffic data.

Traditionally content data has had a high level of privacy protection, and it has been acknowledged that strict privacy requirements for location data other than for traffic data are needed, as they enable exact positioning and hence a permanent surveillance of users.

While it is questionable if the traditional classification of the data in traffic, content and location can be applied to the Internet [4], in contrast to the Directive 97/66/EC, in the EU Directive 2002/58/EC in Art.5 *(Confidentiality of the communication)*, traffic data has been added. Hence, at least according to the new

Directive, traffic data is supposed to have the same level of privacy protection as content data. Thus EU Directive 2002/58/EC is thereby acknowledging that traffic data needs the same level as protection as content data.

The EU Directive 2002/58/EC differentiates between location data other than traffic data, allowing the exact positioning of a mobile user's device, and location data within traffic data, giving geographic information that is often less precise. If used for value-added services, location data other than traffic data has a higher protection. Whereas for traffic data informed consent is required (Art. 6 par. 3,4), for location data other than traffic data either anonymisation or informed consent is required (Art. 9 par.1) with the possibility for users that have given their consent to temporarily refuse the processing for each connection or transmission of a communication (Art.9 par.2).

In this paper, we argue that also traffic data can contain sensitive information about the *relative positioning* and *co-located displacements* of two mobile nodes, and thus also needs a high level of privacy protection.

According to the principle of data minimization and avoidance derived from the principle of necessity of data collection and processing, location data, no matter whether within traffic data or other than traffic data, should be anonymised if the effort involved is reasonable in relation to the desired level of protection. Also for location data within traffic data, users that have given their consent should have the possibility to ”revoke” their consent for each connection or transmission.

C. Privacy-Enhancing Technologies

There are basically two major ways of enhancing privacy in the mobile Internet by technology.

Privacy can be protected most effectively by the first group of privacy technologies that avoid or at least minimize personal data that are exposed on the communication lines and at network sites, and are thus providing anonymity, pseudonymity, unlinkability or unobservability. Mix nets are examples for effective privacy technologies for anonymising communication.

The second way to protect privacy is by using technologies that can control that personal data are only used according to legal provisions. P3P [5] for example, is a technology which can provide technical support for implementing that personal data is only forwarded to web sites with the user's informed consent. According to data protection legislation, informed user consent is often required for the legitimacy of data processing.

1) Mix networks: David Chaum described in [6] a technique based on public key cryptography that allows an electronic mail system to hide who a participant communicates with as well as the content of the communication.

More generally, messages are exchanged through a chain of one or more intermediaries called *“mixes”*. The purpose of a mix is to hide the correspondences between the items in its input and those in its output. The main function of a mix is to: receive and decrypt messages, buffer messages until a defined number of messages has been received, change the sequence of the received messages in a random manner and encrypt and forward the messages to the next mix or to the receiver.

2) Platform for Privacy Preference (P3P) Protocol: The Platform for Privacy Preferences (P3P) Protocol, which has become an official W3C recommendation in April 2002, enables web sites to express their privacy policies in a machine-readable XML format that can be retrieved automatically, interpreted easily and compared with the user's privacy preferences by user agents. Thus, it enables users/ user agents to come to a semi-automated agreement with web sites about the privacy practices for personal data processing by that sites.

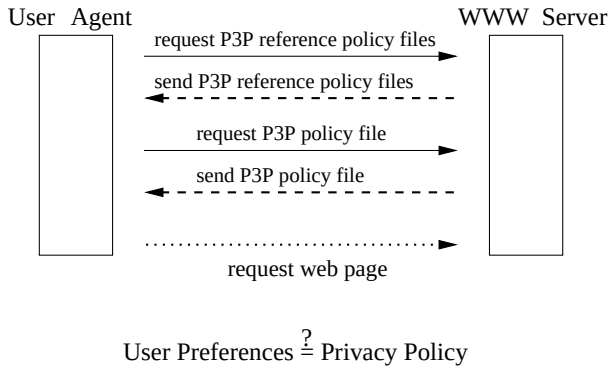

Fig. 2. P3P for informed consent

How a P3P agreement is done is described in [5] and depicted in [Fig. 2]. The P3P user-agent will typically, when an HTTP request is made, fetch a reference file, which is a site map, matching policy files with pages, parts of the site or the whole site, and is typically stored at a well-known location at a website, *”/w3c/p3p.xml”*.

According to this reference file, the appropriate policy file will be retrieved, and matched against the user's preferences. If there is a match, the page will be requested, and if not, the user-agent will take some kind of action to warn the user.

II. Co-located displacements

As explained in [Sect. 1] the home agent of a mobile node keeps track of the binding between the home address (HoA) and the mobile node's care-of-address (CoA) and is fully aware of the network prefix of the link to which a mobile node is attached to. This prefix carries information about the geographical position of the mobile node. Even though a prefix cannot generally be converted into an exact geographical position it will usually confine the possible values of the geographical position to an area that is small in comparison to the area of the surface of the earth.

It also is conceivable that knowledge of the prefix confines the possible positions to a limited area without being able to exactly locate that given area. In that case, however, it will still be possible to determine for two mobile nodes whether or not they are located in the *same* limited area. In other words, for two mobile nodes that use the same home agent, that home agent is aware of possible proximity, and hence the relative positioning, of two mobile nodes.

What is more, however, is that since the care-of address is a function of time, the home agent is able to record at which instance of time a mobile node moves its point of attachment from one foreign link to a new, different, foreign link. That also means that the home agent is able to determine whether or not two mobile nodes move in a co-located fashion. We consider that two mobile nodes that have the same home agent *move in a co-located fashion* if they change to a new care-of-address in the same foreign links a number of times *simultaneously.* An example of such movement is two people traveling in the same car or train. A sketch of how the prefixes of the care-of addresses of two such nodes change as a function of time is given in [Fig. 3]. Note that two nodes generally not change prefix at *exactly* the same time due to the nature of the events that trigger the mobility handovers (signal/noise ratio, network latency etc).

The care-of address as such only determines the geographical position of a mobile node up to the area covered by the foreign link associated with that specific

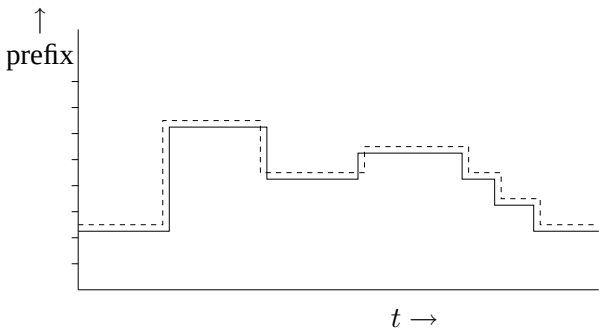

Fig. 3. *Sketch of prefixes of care-of addresses as a function of time. Values, i.e., prefixes, that fall in the same slot on the vertical axis are identical. line:* $CoA^{(i)}(t)$*, dashed line:* $CoA^{(j)}(t)$

care-of address. Movement of a mobile node while connected to the same foreign link will be undetected by the home agent. The care-of addresses of two mobile nodes will allow the home agent to determine the distance between two mobile nodes with an accuracy of approximately the size of the area covered by the foreign link. However, by studying the dynamics of the care-of addresses, it is possible to obtain a more accurate picture of the actual movements of these two mobile nodes. If two mobile nodes change their care-of addresses at almost the same time it is likely that their actual geographical distance was small at the time of change. If this happens a few times in a row it is likely that these nodes were also close at intermediate times. Hence, by studying the *dynamics* of the care-of addresses of two mobile nodes it is possible to extract information about the geographical distance of these two mobile nodes.

A. Analysis of co-located displacements

The care-of address, as a function of time, of a mobile node i will be denoted as $CoA^{(i)}(t)$. For two mobile nodes, i and j, consider the function

$$\Gamma(CoA^{(i)}(t), CoA^{(j)}(t)) = \begin{cases} 1 & \text{prefix } CoA^{(i)}(t) = \text{prefix } CoA^{(j)}(t) \\ 0 & \text{otherwise} \end{cases}$$

Then the integral

$$T = \int_{t_1}^{t_2} dt \Gamma(CoA^{(i)}(t), CoA^{(j)}(t)) \qquad (1)$$

gives the time, within the interval $[t_1, t_2]$, that the nodes i and j were co-located. This could equivalently

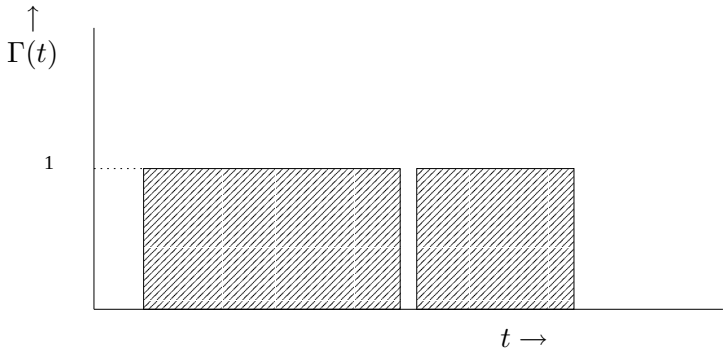

Fig. 4. *The area under $\Gamma(t)$ as a measure of co-location*

be expressed as a percentage $p = T * 100/(t_2 - t_1)$. The measure T does provide information about the *duration* of co-location, irrespective of the *movements* of the nodes i and j. As a consequence two nodes that are connected to the same foreign link during the whole interval $[t_1, t_2]$ will produce $p = 100$, like two nodes that are not static but *move* in a co-located fashion, i.e., simultaneously change to the same new foreign link any number of times. As mentioned above, two roaming nodes will in reality never change prefix at exactly the same instance of time. Such nodes will therefore always produce a number p slightly less than 100 %, even if they move in a fully co-located fashion.

The function $\Gamma(CoA^{(i)}(t), CoA^{(j)}(t))$ implicitly depends on time through the care-of addresses and can also be denoted $\Gamma(t)$. In figure 4 this function is sketched. The shaded area gives the duration of co-location. Small periods of non-co-location that arise when two nodes change prefix only marginally affect the total area, being equivalent to the integral in equation 1.

In order to be able to distinguish co-located roaming from static co-location, i.e. non moving nodes connected to the same link, one can simply count the number of hops where two mobile nodes simultaneously change their care-of address prefixes to the same new value, again within a certain time interval $[t_1, t_2]$. This number, H, is then a functional of the care-of addresses $CoA^{(i)}(t)$, and $CoA^{(j)}(t)$, which in turn are functions of time, and has the following functional form

$$H = h\left[CoA^{(i)}(t), CoA^{(j)}(t), t_1, t_2\right] \qquad (2)$$

The two explicit time arguments t_1 and t_2, indicate the boundaries of the time interval under consideration. H can easily be calculated by analyzing the functions $CoA^{(i)}(t)$ and $CoA^{(j)}(t)$, which in turn can be done by analyzing logged data at, for example, the home agent. Since, as pointed out above, two roaming nodes never change prefix at exactly the same time, it is necessary to use an interval Δt. Two nodes that change prefix of care-of-address at times t and t', where $|t - t'| < \Delta t$ are considered to have changed simultaneously. The interval Δt should be small, at least in comparison to the duration of the entire measurement $t_2 - t_1$.

In summary, the number of simultaneous hops H, of two nodes i and j, can be extracted from logged data by finding all instances where i and j change prefix at times t and t' separated by less than some predetermined amount Δt. Only hops where both the 'old', and the 'new' prefixes are the same should be counted, since otherwise there is no co-location. More formally H can be calculated as follows. If the care-of address $CoA^{(i)}(t)$ changes at times $t_k, k = 1 \cdots n$ then define

$$h_k = \begin{cases} 1 & CoA^{(j)}(t) \text{ shows identical} \\ & \text{change at } t = t', |t' - t_k| < \Delta t \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

Now the quantity H defined in equation 2 can be expressed as the sum

$$H = \sum_{k=1}^{n} h_k \qquad (4)$$

III. Privacy-enhancing technologies for protecting location data

In this section, we will discuss how privacy-enhancing technologies can be applied to technically support and enforce legal privacy requirements of Art. 9 of the EU Directive 2002/58/EC for location data, no matter whether location data other than traffic data or within traffic data.

A. Mix nets for anonymisation of location data

The mix network concept was implemented as part of the Freedom System [7,8]. Freedom is a pseudonymous IP network that provides privacy protection by hiding the user's real IP addresses, email addresses, and other personal identifying information from communication partners and eavesdroppers.

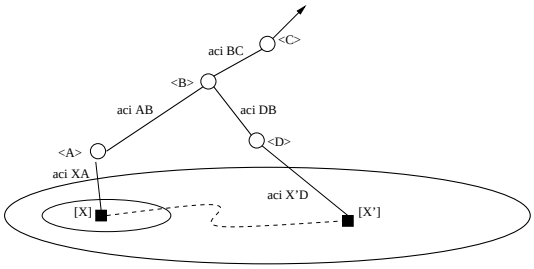

Fig. 5. Mobility extensions for the Freedom System. The virtual circuit is partially recreated during a vertical handover $[X] \rightarrow [X^{'}]$. The exit node $< C >$ is not aware of any mobility.

The Freedom System could be seen as an overlay network composed of globally distributed servers that runs on top of the Internet. Freedom routers or Anonymous Internet Proxies (AIP) are the core network privacy daemons and they are in charge of passing encapsulated packets between themselves until they reach an exit node or AIP wormhole. When a certain AIP runs as an exit node, it works as a traditional network address translator.

Symmetric link encryption is applied between AIP pairs and the freedom-client and the selected AIP entry point to hide the nature and characteristics of the traffic between them. Once the route is created from the freedom client to the wormhole, the data packets travel toward the wormhole over the virtual circuit, being link decrypted, telescope unwrapped and finally link encrypted at each point. The data is routed to the next hop by use of an Anonymous Circuit Identifier mapping table. The ACIs indicate, along with a packet's implicit source address and port, the next hop in a particular route.

When a freedom client communicates with a correspondent node via a previously built virtual circuit in the Freedom System, the correspondent node sees that the traffic as coming from the wormhole IP address instead of the client's real IP address.

In [9] we introduced a set of protocol extensions to the Freedom System architecture to permit a mobile node to seamlessly roam among IP subnetworks and media types while remaining untraceable and pseudonymous. The extensions make it possible to support transparency above the IP layer, including the maintenance of active connections in the same way that MobileIP does but with the addition that the home and foreign network are unlinkable [Fig. 5].

The concept of a mix network for location based services was also introduced in [10] where a Privacy Enhanced-Location Based Services (PE-LBS) proxy can be configured to act as a "mix" by buffering and changing the sequence of the service requests. The mobile device can use a chain of PE-LBS proxies configured as a "mixing network" to forward a location based service request. The architecture allows a mobile node to request location based services via a mix-network hiding the network location of the mobile device while providing service accountability.

B. P3P and processing of location data in MobileIP

The Platform for Privacy Preferences (P3P) Protocol can be used as a technical means for technically supporting the privacy principle of informed consent, and also for allowing users to later revoke their consent. Although P3P is a standard for controlling the personal data processing by web servers, we will discuss how it could also be used for obtaining informed consent for the processing of location data within traffic data by home agents for value-added services, such as location based services.

In the future, more effective solutions for automated privacy agreements between mobile nodes and home agents could be based on compact privacy policy or preference information included in newly to be defined extension headers for Mobile IPv6. Such a solution, however, will first require new protocol extensions, whereas a solution based on P3P can easily be implemented already today with available technologies.

Often there is a close administrative relationship between the owner of a mobile node and the owner of that mobile node's home agent. For example, a company that provides mobile nodes for its employees is also operating home agents for those mobile nodes, or a home agent could even be operated by the mobile user. If there is a trust relation between the mobile user and the owner of the home agent, an agreement about data processing practices do not have to be automated but can as well be done off-line. However, home agents could also be owned by a service provider or other organizations to which no close trust relation exists. Besides, if we demand the same level of privacy protection for location data within traffic data as for location data other than traffic data, mobile users should still have the possibility to revoke their consent for the

processing of location information within traffic data for each connection or transmission of a communication (as required by Art. 9 par. 2 for location data other than traffic data). Also for the enforcement of this requirement, an online solution as provided by P3P is required.

For enabling P3P agreements a home agent needs to have a web server interface and has to have a P3P privacy policy containing a statement describing data practices that are applied to location data.

```
<STATEMENT>
 <PURPOSE>
     <current/>
     <other-purposes required = "opt-in">
     Location-based-Services
     </other-purpose>
 </PURPOSE>
  <RECIPIENT>
              <ours/>
         </RECIPIENT>
         <RETENTION>
    <no-retention/>
       </RETENTION>
         <DATAGROUP>
             <DATA> ref="#dynamic.miscdata">
                    <CATEGORIES>
                          <location/>
                     </CATEGORIES>
                </DATA>
        </DATAGROUP>
</STATEMENT>
```

Fig. 6. P3P policy statement for a EU Directive 2002/58/EC compliant processing of location data.

In order to be compliant with the EU Directive 2002/58/EC, location data within traffic data should only be processed for the transmission of a communication (Art. 6 par. 1) or for marketing its own electronic communication services or for the provision of value-added services, such as location-based services (Art. 6 par.3). [Fig. 6] shows an example P3P policy statement for location data, allowing its processing for the current purpose of transmitting a communication and for location-based services. The opt-in requirement should be used to state that location data can only be processed for location-based services if the user explicitly requests that service and thus gives his/her consent for the use of location data for location-based services. At a policy's "opturi" link, instructions are provided for users how to decline from their request. Hence, the home agent could set up a web site that allows mobile users to fill-in forms for granting or revoking their consent for the processing of location data for the specified value-added services. This guarantees that also Art.9 par.2 can be technically supported.

Within the P3P policy statement, the RECIPIENT element should be set to <ours/>, meaning that the location data is only handled by the Home Agent or possibly entities processing the data on its behalf for the completion of the value-added service, as required by Art. 6 par.4 and Art. 9 par.3.

The RETENTION element that indicates the kind of retention policy that applies to the data should be set to <no-retention/> or <stated-purpose/> to state that the data are only processed for the duration necessary for the value-added service as required by Art. 6 par.3 and Art.9 par.1.

The mobile node has to have a P3P-compliant user agent including P3P privacy preferences defined by its user for the processing of location data. By accessing the Home Agent's web site, the mobile user can check the Home agent's privacy policy for processing of location data and can fill-in a form for requesting a value-added service, and for thereby giving his/her informed consent for the processing of traffic data for that service. A mobile user should evaluate the home agent's privacy policy at the time that she/he chooses a mobile node, and should reevaluate it before the expiry period of the policy file has passed, or in case that the mobile user has changed his/her preferences or wants to revoke his/her consent.

A problem, however, is that location information is included in all messages sent by a mobile node to its home agent. Thus, when the P3P user agent of a mobile node is fetching the P3P reference file and the policy file, it is already transferring location information with those requests, even though there has not been a successful P3P agreement with that agent yet. The home agent's web site should hence follow the so-called safe-zone practices for communications which take place as part of fetching a P3P policy or policy reference file, and thus should not collect location information that is available within the safe zone. If a user does not want to rely on the safe-zone practices, she/he should preferably initiate P3P negotiations at times that her/his node is located in its home network. If a mobile user does not succeed to select a home agent that fulfills her/his privacy preferences, she/he should

have the option to use anonymous communication.

P3P has been criticized by the Art.29 Data Protection Working Party [11,12] and others, as it cannot in itself secure privacy on the Web. Hence it needs to be applied according to a regulatory framework, such as given by EU Directive 2002/58/EC. Besides, P3P cannot ensure that web sites really follow privacy policies as they claim to do. Third party monitoring can enhance control over the compliance with the privacy policies published at web sites.

IV. Conclusions

In this paper, we have shown that traffic data in MobileIP-based networks can also contain sensitive information about the relative positioning and co-located displacements of two mobile nodes, and thus also needs high level of privacy protection. By studying the dynamics of the care-of addresses of two mobile nodes it is possible to extract information about the geographical distance of these two mobile nodes. The co-located displacements in MobileIP allow to the home agent to determine whether or not a set of mobile nodes move in co-located fashion.

Privacy-enhancing technologies should be applied to technically enforce legal privacy requirements of Art. 9 of the EU Directive 2002/58/EC for location data, no matter whether location data other than traffic data or within traffic data.

According to the privacy principle of data minimization and data avoidance, location data should be anonymized if the effect involved is reasonable in relation to the desired effect. Mix-nets based architectures (as presented in section III-A) provide an effective means for anonymising location data, and should preferably be provided to mobile users in order to fulfill the requirements of Art. 9 par. 1. We have also shown how the Platform for Privacy Preferences (P3P) Protocol can be used as an Online mechanism for obtaining informed consent of mobile users for the use of location data for value-added services, and also for allowing users to later revoke their consent, as required by Art. 9 par.1 and par.2. Hence, we have shown how existing technologies can be used to provide different levels of location privacy, even though we strongly propose that the future developments in the next generation Internet Protocol should also directly include features for location privacy.

References

[1] C. Perkins, IP Mobility Support, RFC 2002, October 1996.

[2] Directive 2002/58/EC of the European Parliament and of the Council concerning the processing of personal data and the protection of privacy in the electronic communications sector, Brussels, 12 July 2002, http://www.etsi.org/public-interest/Documents/Directives/Standardization/Data_Privacy_Directive.pdf

[3] European Union, Directive 97/66/EC of the European Parliament and of the Council Concerning the Processing of Personal Data and the Protection of Privacy in the Telecommunications Sector of 15 December 1997. http://europa.eu.int/ISPO/infosoc/telecompolicy/en/9766en.pdf

[4] A. Escudero, Location Privacy in mobile Internet in the context of the European Union Data Protection Policy. Proceedings of INET2002. Washington DC. June 2002.

[5] The Platform for Privacy Preferences 1.0 (P3P1.0) Specification , W3C Recommendation, 16 April 2002, http://www.w3.org/TR/P3P/

[6] D. Chaum, Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. Communications of the ACM (24)2, 1981.

[7] I. Goldberg, and A. Shostack, Freedom Network 1.0 Architecture and Protocols. 1999.

[8] A. Escudero, M. Hedenfalk, and P. Heselius, Flying Freedom: Location Privacy in Mobile Internetworking. INET2001. Stockholm. June 2001.

[9] A. Escudero, Anonymous and Untraceable Communications: Location Privacy in Mobile Internetworking. Licentiate Thesis ISSN 1403-5288. May 2001.

[10] A. Escudero, and G. Q. Maguire Jr, Role(s) of proxy in Location Based Services. Proceedings of 13th IEEE International Symposium IEEE on Personal, Indoors and Mobile Radio Communications, Vol.3 pp 1252-1257, Lisbon. September 2002.

[11] Working Party on the Protection of Individuals with regard to processing of Personal Data, Opinion 1/98, Platform for Privacy Preferences (P3P) and the Open Profiling Standard (OPS), adopted on 16 June 1998, http://europa.eu.int/comm/internal_market/en/dataprot/wpdocs/wp11en.htm

[12] Article 29 - Data Protection Working Party, "Privacy on the Internet - An integrated EU approach to Online Data Protection", adopted on 21st November 2000, http://europa.eu.int/comm/internal_market/en/dataprot/wpdocs/wp37en.pdf

APPENDIX A

Article 29 Data Protection Working Party

"Opinion 2/2002 on the use of unique identifiers in telecommunication terminal equipments: the example of IPv6" [1]

http://europa.eu.int/comm/internal_market/en/dataprot/wpdocs/wp58_en.pdf
May 2002

[1] The document is reproduced in the original format

10750/02/EN/Final
WP 58

Opinion 2/2002
on the use of unique identifiers in telecommunication terminal equipments: the example of IPv6

Adopted on 30 May 2002

set up by Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995[1],

having regard to Articles 29 and 30 paragraphs 1 (a) and 3 of that Directive,

having regard to its Rules of Procedure and in particular to Articles 12 and 14 thereof,

has adopted the present Opinion:

The Commission Communication on IPv6

On 21 February 2002, the European Commission adopted a communication to the Council and the European Parliament, focusing on the next generation Internet and the priorities for action in migrating to the new Internet protocol IPv6. This communication takes place in the context of the current development of network services and terminal communication equipment enabled to connect to the network.

The new Internet protocol has been elaborated with a view to facilitate and harmonise the possibilities of connection to the network using multiple terminal equipment, such as mobile phones, personal computers or personal digital assistants, using wireless or cable facilities.

While these developments can only be encouraged, the Working Party would like to stress the need for a careful and in-depth study of the implications of the new protocol in terms of protection of personal data.

The Working Party welcomes the position taken by the Commission in its communication, according to which privacy issues are to be considered in the further development of the Internet. The Working Party stresses, however, that privacy issues raised by the development of the new protocol IPv6 have not been solved yet.

In particular, the possibility of the integration of an unique identification number in the IP address as designed according to the new protocol raises specific concern. In this respect, the Working party regrets that it has not been consulted prior to the adoption of the communication and it expresses the wish to be involved in the coming works taking place on IPv6 at European level.

Data Protection aspects related to the use of unique identifiers in telecommunication terminal equipment

The Working Party takes note of the fact that the International Working Group on data protection in telecommunications has recently issued a working paper on the question of the use of unique identifiers in telecommunication terminal equipment, and it would like to thank the Working group for the work achieved on that subject.

[1] Official Journal no. L 281 of 23/11/1995, p. 31, available at: http://europa.eu.int/comm/internal_market/en/dataprot/index.htm

The Working Party endorses the conclusions of the working paper adopted in Auckland on 27 March 2002[2], and would like to support its findings by recalling in particular the application of several principles explicitly mentioned in EU directive 95/46 concerning the protection of personal data and the free movement of such data, and EU Directive 97/66 concerning the protection of personal data in the telecommunication sector[3].

The Working Party wishes to emphasise that IP addresses attributed to Internet users are personal data[4] and are protected by EU Directives 95/46 and 97/66.

With reference to the work already achieved with regard to the protection of personal data on the Internet[5], the Working Party would like to stress specifically the following points:

- The unique identifier of an interface, such as the one that might be integrated in IPv6, would constitute an identifier of general application and its use is regulated as such in the legislation of the member States of the EU.

- The principle of proportionality implies that, making a balance between the fundamental rights of data subjects and the interests of different actors involved in the transmission of telecommunication data (such as companies, telecommunication access providers), as few personal data as possible have to be processed.

  This principle has implications on the one hand on the design of the new communication protocols and devices, and on the other hand on the content of national policies related to the processing of telecommunication data: while technology is *per se* neutral, applications and design of new telecommunication devices should be privacy compliant by default. Besides, it should be avoided to generalise measures forcing the systematic identifiable character of telecommunication data.

In that perspective, in the framework of a telecommunication connection, network and access providers should offer to any user the option to use the network or to access the services anonymously or using a pseudonym.

---

[2] See the annex to this document.

[3] Directive 97/66 is being amended in order to take into account technological developments. The provisions of the new directive are intended to protect users of publicly available electronic communications services, regardless of the technologies used.

[4] As recital 26 of Directive 95/46 specifies, data are qualified as personal data as soon as a link can be established with the identity of the data subject (in this case, the user of the IP address) by the controller or any person using reasonable means. In the case of IP addresses the ISP is always able to make a link between the user identity and the IP addresses and so may be other parties, for instance by making use of available registers of allocated IP addresses or by using other existing technical means.

[5]
- Working document: Processing of Personal Data on the Internet, adopted by the Working Party on 23 February 1999, WP 16, 5013/99/EN/final;
- Recommendation 1/99 on Invisible and Automatic Processing of Personal Data on the Internet Performed by Software and Hardware, adopted by the Working Party on 23 February 1999, 5093/98/EN/final, WP 17;
- Recommendation 2/99 on the respect of privacy in the context of interception of telecommunications, adopted on 3 May 1999, 5005/99/final, WP 18;
- Recommendation 3/99 on the preservation of traffic data by *Internet Service Provide*rs for law enforcement purposes, adopted on 7 September 1999, 5085/99/EN/final, WP 25;
- Opinion 1/2000 on certain data protection aspects of electronic commerce, Presented by the Internet Task Force, adopted on 3 February 2000, 5007/00/EN/final, WP 28;
- Opinion 2/2000 concerning the general review of the telecommunications legal framework, presented by the Internet Task Force, adopted on 3 February 2000, WP 29, 5009/00/EN/final;
- Opinion 7/2000 on the European Commission Proposal for a directive of the European Parliament and the Council concerning the processing of personal data and the protection of privacy in the telecommunications sector of 12 July 2000 COM (2000) 385, adopted on 2 November 2000, WP 36.

EC Directive 97/66 provides for the possibility for any user to restrict the identification of calling and connected addresses. In Internet communications, anonymity could be reached using solutions such as regularly changing IP addresses used by an individual[6].

- Considering the risks of manipulation and fraudulent use of a unique identifier, the working party recalls that protection measures are needed, taking into account in particular that telecommunication providers are responsible for the security of services they offer. In the framework of the European Union legislation, access providers are obliged to inform subscribers of residual security risks.

- The requirements for privacy compliant default settings in communication devices and for privacy compliance of telecommunication services have been implemented at European level through specific obligations lying mainly on producers of telecommunications equipment, and on telecommunication operators and service providers[7].

Conclusion

The working group strongly encourages research initiatives having as purpose the elaboration of technical solutions to protect the privacy of telecommunication data.

The Working Party is aware that initiatives have already been taken in different working groups in order to find technical solutions to some identified privacy risks, and it considers it necessary to enter into a dialogue in particular with representatives of these groups, and in particular, the Internet Engineering Task Force and the IPv6 Task Force.

The Working Party reserves the possibility to take further steps while evaluating the new design of communication protocols, products and services and while continuing dialogue with actors involved in the design of these new communication tools.

---

[6] Such solution has already been adopted by some access providers, who change approximately every two days the IP address of their ADSL clients.
The implementation of some terminal equipment already takes into account the orientations of RFC 3041 of the Internet Engineering Task Force (IETF), "privacy extensions for stateless address autoconfiguration in Ipv6", January 2001: the terminal equipment uses two types of addresses : an address is generated based on the unique MAC address, and is used for entering communications (e.g. the terminal is always reachable using that permanent address), and another address generated on a (pseudo) random basis, to be used at the initiative of the terminal for outgoing connections.
Thus, when the terminal (and the user behind) is responsible for the connection, it could not be identified through its MAC address.

[7] See Directive 97/66 on the protection of privacy in the telecommunications sector, and Directive 99/5 on radio equipment and telecommunications terminal equipment and the mutual recognition of their conformity, Official Journal L 091, 07/04/1999.

Annex

Working paper on the use of unique identifiers in telecommunication terminal equipments: the example of IPv6

31st meeting of the International Working Group on Data Protection in Telecommunications on 26-27 March 2002 in Auckland (New Zealand)

Due to a forseeable shortage in the protocol used today for most of the Internet connections (IP version 4), a change of design in the protocol has been elaborated by the international Internet Engineering Task Force (IETF). This new protocol, IPv6, uses a string of 128 bits instead of 32 bits in the former version, to constitute each individual IP address on the Internet.

This new address, thanks to its enlarged capacities, presents many advantages and enables new facilities such as multicasting (quicker transmission of large amounts of data to multiple recipients, e.g. video on-line), voice over IP, etc.

However, the new protocol also raises concerns, as it has been designed in such a way that each IP address can be partly constituted of a unique serie of numbers like a global unique identifier. The introduction of IPv6 might lead to increased risks of profiling of user activities on the Internet[8].

The following preliminary considerations identify the risks and recall the privacy principles to take into consideration while using a unique identifier in the constitution of IP addresses.

I. Identified risks

The characteristics of IPv6 lead to the identification of specific privacy risks, which will depend on the configuration of the new protocol.

- *Profiling issues* are at stake if a unique identifier (the interface identifier e.g. based on the unique MAC address of the ethernet card) is integrated in the IP address of each electronic communication device of the user. In such case, all communications of the user can be linked together, much easier than using cookies as they exist today.

- ***security and confidentiality issues*** **can be identified. These risks are linked with the development of network services, which implies multiplication of the type of terminals connected to the network using the same communication protocol: mobile phones, personal computers, electronic agents controlling home devices (heating, light, alarms, etc.).**

  The new IPv6 protocol allows stable connections, with maintenance of the same address, even when a terminal is moving on the network. Security and

[8] Overall profiling of activities of a user might even be feasible when the same terminal equipment is used in different networks.

confidentiality aspects are at stake here, as there is a risk of identification of location data of this mobile node[9].

II. Data protection principles applicable to IPv6

The working group deems it necessary to draw the attention of all the actors responsible in the elaboration and the implementation of the new protocol, about the national and international legal requirements governing privacy and security of telecommunications.

It is now widely recognised that IP address - and *a fortiori* a unique identification number integrated in the address – can be considered as personal data in the sense of the legal framework[10].

In line with his previous work and the common positions already adopted on that subject[11], the Working Group recalls the following principles, which should be taken into account while implementing the new Internet protocol.

Telecommunications infrastructure and technical devices have to be designed in a way that either no personal data at all or as few personal data as technically possible are used to run networks and services. The unique identifier of an interface as integrated in IPv6 would constitute an identifier of general application.

§ In contradiction with the principle of data minimisation, such use of a unique identifier constitutes a risk of profiling of individuals for all their activities in connection with a network.

§ The protection of the fundamental right to privacy against such risk of profiling must prevail while analysing the different aspects of the new protocol, such as its facility of management.

§ Traffic data, and in particular location data, deserve a specific protection considering their sensitive character[12].
If location information has to be generated in the framework of the use of mobile devices and other objects connected via IP, such information must be protected against unlawful interception and misuse. It should also be avoided that the location information (and the changing in this location information depending on the

---

[9] See e.g. A. Escudero Pascual, "Anonymous and untraceable communications: location privacy in mobile iternetworking", 16 May 2001; "Location privacy in Ipv6 – Tracking the binding updates", 31 August 2001; http://www.it.kth.se/~aep/

[10] See e.g. at European level, the Communication of the Commission on the Organisation and Management of the Internet Domain Name System of April 2000, and the documents adopted by the Art. 29 Data Protection Working Party, in particular "Privacy on the Internet - An integrated EU Approach to On-line Data Protection", WP 37, 21 November 2000.

[11] Common Position regarding Online Profiles on the Internet, adopted at the 27th meeting of the Working Group on 4/5 May 2000;
§ Common Position on Privacy and location information in mobile communications services, adopted at the 29th meeting of the Working Group on 15/16 February 2001;
§ Ten Commandments to protect Privacy in the Internet World
Common Position on Incorporation of telecommunications-specific principles in multilateral privacy agreements, adopted at the 28th meeting of the Working Group on 13/14 September 2000.
http://www.datenschutz-berlin.de/doc/int/iwgdpt/inter_en.htm

[12] See the Common Position on Privacy and location information in mobile communications services, adopted at the 29th meeting of the Working Group on 15/16 February 2001.

.

movement of the mobile user), is transmitted non-encrypted to the recipient of the information via the header of the IP address used.

Protocols, products and services should be designed to offer choices for permanent or volatile addresses. The default settings should be on a high level of privacy protection.

Since these protocols, products and services are continuously evolving, the Working Group will have to monitor closely the developments and to call for specific regulation if necessary.

Done at Brussels, 30 May 2002

For the Working Party

The Chairman

Stefano RODOTA

APPENDIX B

Alberto Escudero-Pascual

"Location data is as sensitive as content data" [2]

Contribution to the European Union Forum on Cybercrime
http://cybercrime-forum.jrc.it
Brussels, 27th November 2001

[2] The document is reproduced in the original format

Location data is as sensitive as content data

Contribution to the European Union Forum on Cybercrime

Alberto Escudero Pascual <alberto@it.kth.se>
Royal Institute of Technology - Sweden

2001/11/27

Abstract

The following contribution to the EU Forum on cybercrime (27th Nov 2001) is based on the results of our research at the Royal Institute of Technology concerning location data in mobile networks. The information included in the Internet Protocol headers plus the mobile terminal locations can determine - with high precision - our human interactions, interests and behavioural habits. Therefore, our main statement is that 'location data' should be considered to be just as sensitive as 'content data' due to the categories of information that can be extracted from location data sets.

1 Background

Wireless internet access of the Kista - IT University (KTH) network began in October 1999 as a research project in the Telecommunication Systems Lab *(TSLab)* at the former Department of Teleinformatics (Sweden).

The main research objective was to study the possibility of adopting wireless access as part of the IT University network infrastructure and to evaluate the level of security required by networks of this kind.

Based on our previous experience running smaller scale wireless networks, a new innovative networking environment was conceived to offer Internet access for the five hundred students and researchers of the IT-University study programme. Each student or researcher uses an IEEE 802.11b compliant PCMCIA card to get wireless connectivity using a set of access points available in three different buildings and common areas. Mobility is supported between the radio cells using link level handover or MobileIPv4 to roam between IP networks.

The wireless access allows students and researchers to attend lectures with their laptops, take notes and see online documentation while their mail arrives in their laptop mailboxes.

The IT University/KTH wireless environment provides an unvaluable research environment to study new privacy concerns related to location data before the massive deployment of the 3G networks.

2 Personally Identifiable Information in mobile internetworking

The protection of privacy in computer networks emerged as a research topic for the TeleSystems Laboratory at KTH in Sweden in mid-June 2000. Data protection and privacy is rapidly becoming one of the most important issues on the Internet today. More and more Internet sites are collecting personal information from users through forms, cookies, online registrations, or surveys than ever before.

New commercial services are springing up that can exploit the ability of mobile communication service providers to determine the geographic location of their users. The new wireless technologies offer mobility; at the same time they offer *"location information"* that is being used to provide new "location-aware services".

Our general goal is to study if it is possible to provide a mobile user with internet communication while simultaneously protecting personally identifiable information. More specifically: we are concerned with the privacy aspects of location information.

By *location privacy in mobile internetworking* we mean the capability of a mobile node to conceal geographical information from third parties while the user is on the move. We intend to address the privacy concerns that these technologies raise and explore different solutions.

2.1 The "storebror" Big Brother System.

The *Big Brother System* [5] was built when the Kista - IT University wireless network was being designed in October 2000. Initially designed as a networking tool to help us with the positioning of the wireless access points. Big brother is a monitoring system that detects the movements of the wireless users at the Kista IT-University.

Every 60 seconds `orwell.it.kth.se` updates the position of each of the mobile nodes. This position is stored in a private database in where personal data is protected by providing a *pseudonym* for each user in the network.

Use of the "*Storebror System*", which is a well-known and trusted system for monitoring wireless positioning, raised lots of concerns among our academic community [2],[3]. Questions raised include: How and who should handle this kind information? Which technical and legal means needs to be deployed to protect personally identifiable information in mobile internetworking?

3 Conclusions

Location data sets provide by themselves without aditional content data with enough information to draw the map of human relationships. When it comes to location information is very difficult to draw a line between traffic data and content data.

By agreggation and correlation of location data sets it is possible to determine which places are visited more often and the human afinity of their visitors. Location data records contain the information of “where, when, how long and with whom” and that information should be consider **as sensitive as the content data**.

References

[1] Escudero A, Pehrson B, Pelletta E, Vatn J.O, Wiatr P. *"Wireless access in Kista - IT University: MobileIPv4 integration in a IEEE 802.11b”*. 11th IEEE Workshop on Local and Metropolitan Area Networks. LANMAN2001. Co. USA. March 2001.

[2] Andersson. S. NyTeknik. *“På KTH utvecklas teknik att stoppa övervakning*”. Nov 2000
`http://www.nyteknik.se/pub/pub26_3.asp?art_id=12932`

[3] Aftonbladet Nyheter *“KTH-studenter övervakas via trådlöst nätverk Storebror ser dig”*.Nov 2000
`http://www.aftonbladet.se/nyheter/0011/03/kth.html`

[4] Escudero A, Hedenfalk, M. Heselius P. *“Location Privacy in Mobile Internet - An extension to Freedom Network"*. April 2001

[5] *Storebror “Big Brother” System.* October 2000.
`http://www.flyinglinux.net/bigbrother`

TRITA-IMIT-TSLAB AVH 02:01
ISSN 1651-4114
ISRN KTH/IMIT/TSLAB/AVH-02/01--SE

Telecommunication Systems Laboratory
Department of Microelectronics and
Information Technology
Royal Institute of Technology
Electrum 204
SE-164 40 Kista
Sweden