Network Working Group
Internet-Draft
Intended status: Informational
Expires: September 15, 2011

B. Aboba
Microsoft Corporation
J. Morris
CDT
J. Peterson
NeuStar, Inc.
H. Tschofenig
Nokia Siemens Networks

March 14, 2011

Privacy Considerations for Internet Protocols
draft-morris-privacy-considerations-03.txt

Abstract

This document aims to make protocol designers aware of privacy-related design choices and offers guidance for developing privacy considerations for IETF documents. While specifications cannot police the implementation community, nonetheless protocol architects must play in the improvement of privacy, both by making a conscious decision to design for privacy, and by documenting privacy risks in protocol designs.

This document is discussed on the Internet Privacy Discussion mailing list (see https://www.ietf.org/mailman/listinfo/ietf-privacy).

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 15, 2011.

Table of Contents

1. Introduction 3
2. Scope 5
3. Threat Model 9
4. Guidelines 11
5. Example 13
5.1. Presence 13
5.2. AAA for Network Access 15
6. Security Considerations 18
7. IANA Considerations 19
8. Acknowledgements 20
9. References 21
9.1. Normative References 21
9.2. Informative References 21
Appendix A. Historical Background 25
Authors' Addresses 29

1. Introduction

The IETF produces specifications that aim to make the Internet better. Those specifications fall into a number of different categories, including protocol specifications, best current practice descriptions, and architectural documentations. While IETF documents are typically implementation-agnostic, they are often, if not always, impacted by fundamental architectural design decisions. These decision decisions in turn hinge on technical aspects, predictions about deployment incentives, operational considerations, regulatory concerns, security frameworks, and so on.

This document aims to make protocol designers aware of privacy-related design choices and offers guidance for developing privacy considerations for IETF documents. While specifications cannot police the implementation community, nonetheless protocol architects must play in the improvement of privacy, both by making a conscious decision to design for privacy, and by documenting privacy risks in protocol designs. While discuss the limitations of standards activities in Section 2, we maintain that the IETF community in its mandate to "make the Internet better" has a role to play in making its specifications, and the Internet, more privacy friendly. This must spring from awareness of how design decisions impact privacy, and must be reflected in both protocol design and in the documentation of potential privacy challenges in the deployment a single protocol or an entire suite of protocols.

From the activities in the industry, one can observe three schools of thought in the work on privacy, namely

Privacy by Technology:

> This approach considers the assurance of privacy in the design of a protocol as a technical problem. For example, the design of a specific application may heighten privacy by sharing fewer data items with other parties (i.e. data minimization). Limiting data sharing also avoids the need for evaluation on how data-related consent is obtained, to define policies around how to protect data, etc. Ultimately, different architectural designs will lead to different results with respect to privacy.
>
> Examples in this area of location privacy can be found in [EFF-Privacy]. These solution often make heavy use of cryptographic techniques, such as threshold cryptography and secret sharing schemes.

Privacy by Policy:

In this approach, privacy protection happens through establishing the consent of the user to a set of privacy policies. Hence, protection of the user privacy is largely the responsibility of the company collecting, processing, and storing personal data. Notices and choices are offered to the customer and backed-up by an appropriate legal framework.

An example of this approach for the privacy of location-based services is the recent publication by CTIA [CTIA].

Policy/Technology Hybrid:

This approach targets a middle-ground where some privacy-enhancing features can be provided by technology, and made attractive to implementers (via explicit best current practices for implementation, configuration and deployment, or by raising awareness implicitly via a discussion about privacy in technical specifications) but other aspects can only be provided and enforced by the parties who control the deployment. Deployments often base their decisions on the existence of a plausible legal framework.

The authors believe that the policy/technology hybrid approach is the most practical for engineers in the IETF.

This remainder of this document is structured as follows: In Section 2, we illustrate what is in scope of the IETF and where the responsibility of the IETF ends. In Section 3, we discuss the main threat model for privacy considerations. In Section 4, we propose guidelines for documenting privacy within IETF specifications, and in Section 5 we examine the privacy characteristics of a few exemplary IETF protocols and explain what privacy features have been provided to date. In the appendix we provide a brief introduction to the concept of privacy in Appendix A.

2. Scope

The IETF at large produces specifications that typically fall into the following categories:

- Process specifications (e.g. WG shepherding guidelines described in RFC 4858 [RFC4858]) These documents aim to document and to improve the work style within the IETF.

- Building blocks (e.g. cryptographic algorithms, MIME types registrations). These specifications are meant to be used with other protocols one or several communication paradigms.

- Architectural descriptions (for example, on IP-based emergency services [I-D.ietf-ecrit-framework], Internet Mail [RFC5598])

- Best current practices (e.g. Guidance for Authentication, Authorization, and Accounting (AAA) Key Management [RFC4962])

- Policy statements (e.g. IETF Policy on Wiretapping [RFC2804])

Often, the architectural description is compiled some time after the deployment has long been ongoing and therefore those who implement and those who deploy have to make their own determination of which protocols they would like to glue together to a complete system. This type of work style has the advantage that protocol designers are encouraged to write their specifications in a flexible way so that they can be used in multiple contexts with different deployment scenarios without a huge amount of interdependency between the components. [Tussle] highlights the importance of such an approach and [I-D.morris-policy-cons] offers a more detailed discussion.

This work style has an important consequence for the scope of privacy work in the IETF, namely

- the standardization work focuses on those parts where interoperability is really essentially rather than describing a specific instantiation of an architecture and therefore leaving a lot of choices for deployments.

- application internal functionality, such as API, and details about databases are outside the scope of the IETF

- regulatory requirements of different juristictions are not part of the IETF work either.

Here is an example that aims to illustrate the boundaries of the IETF work: Imagine a social networking site that allows user registration,

requires user authentication prior to usage, and offers its functionality for Web browser users via HTTP, real-time messaging functionality via XMPP, and email notifications. Additionally, support for data sharing with other Internet service providers is provided by OAuth.

While HTTP, XMPP, Email, and OAuth are IETF specifications they only define how the protocol behavior on the wire looks like. They certainly have an architectural spirit that has enormous impact on the protocol mechanisms and the set of specifications that are required. However, IETF specifications would not go into details of how the user has to register, what type of data he has to provide to this social networking site, how long transaction data is kept, how requirements for lawful intercept are met, how authorization policies are designed to let users know more about data they share with other Internet services, how the user's data is secured against authorized access, whether the HTTP communication exchange between the browser and the social networking site is using TLS or not, what data is uploaded by the user, how the privacy policy of the social networking site should look like, etc.

Another example is the usage of HTTP for the Web. HTTP is published in RFC 2616 and was designed to allow the exchange of arbitrary data. An analysis of potential privacy problems would consider what type of data is exchanged, how this data is stored and processed. Hence, the analysis for a static webpage by a company would different than the usage of HTTP for exchanging health records. A protocol designer working on HTTP extensions (such as WebDAV) it would therefore be difficult to describe all possible privacy considersations given that the space of possible usage is essentially unlimited.

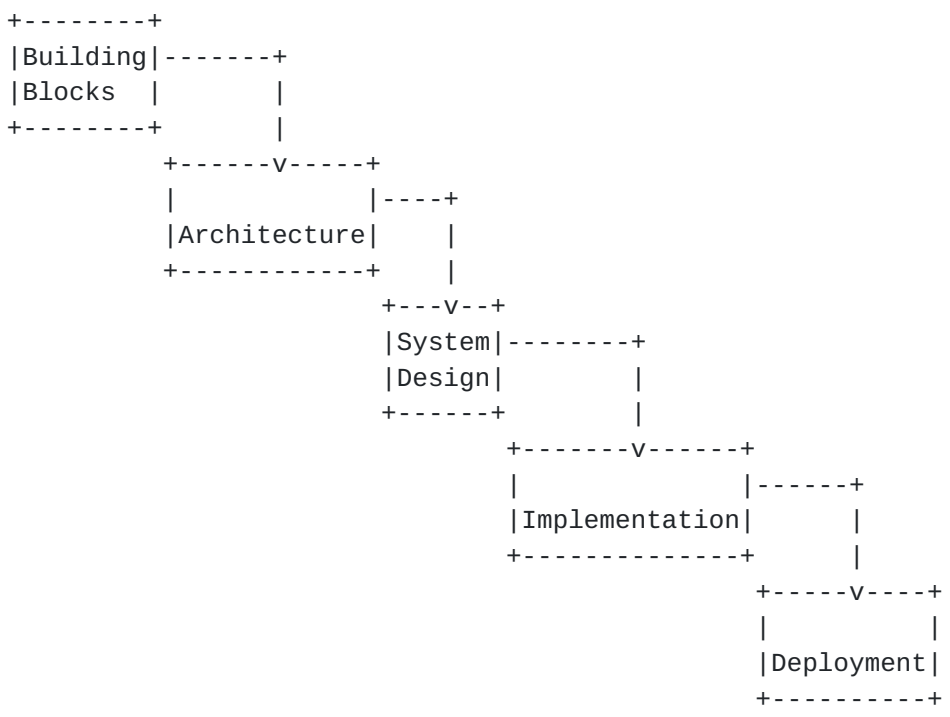


Figure 1: Development Process

Figure 1 shows a typical development process. IETF work often starts with identifying building blocks that ccan then be used in different architectural variants useful for a wide range of usage scenarios. Before implementation activities start a software architecture needs to evaluable which components to integrate, how to provide proper performance characteristics, etc. Finally, the implemented work needs to be deployed. Privacy considerations play a role along the entire process.

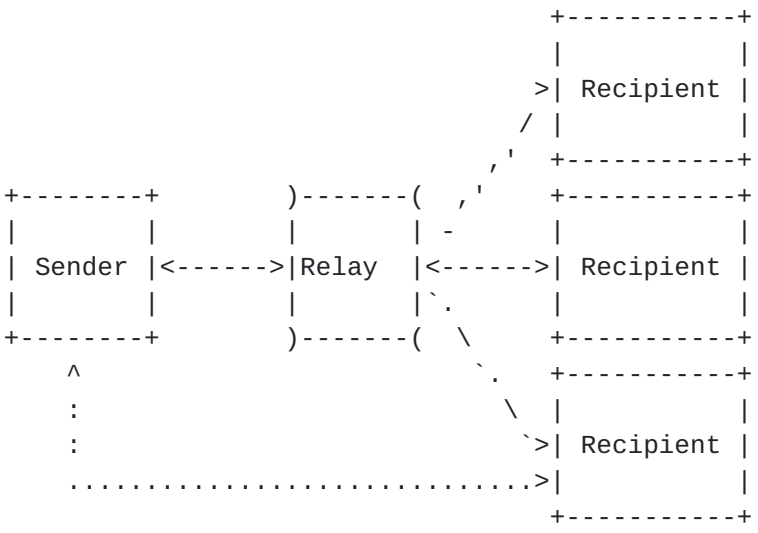
> To pick an example from the security field consider the NIST Framework for Designing Cryptographic Key Management Systems [SP800-130], NIST SP 800-130. SP 800-130 provides a number of recommendations that can be addressed largely during the system design phase as well as in the implementation phase of product development. The cryptographic building blocks and the underlying architecture is assumed to be sound. Even with well-design cryptographic components there are plenty of possibilities to introduce security vulnerabilities in the later stage of the development cycle.

Similiar to the work on security the impact of work in standards developing organizations is limited. Neverthelesss, discussing potential privacy problems and considering privacy in the design of an IETF protocol can offer system architects and those deploying systems additional insights. The rest of this document is focused on

illustrating how protocol designers can consider privacy in their design decisions, as they do factors like security, congestion control, scalability, operations and management, etc.

3. Threat Model

To consider privacy in protocol design it useful to think about the overall communication architecture and what the different actors could do. This analysis is similar to a threat analysis found in security consideration sections of IETF documents. See also RFC 4101 [RFC4101] for an illustration on how to write protocol models. In Figure 2 we show a communication model found in many of today's protocols where a sender wants to establish communication with some recipient and thereby uses some form of intermediary (referred as relay in Figure 2. In some cases this intermediary stays in the communication path for the entire duration of the communication and sometimes it is only used for communication establishment, for either inbound or outbound communication. In rare cases they may even be a series of relays that are traversed.



Legend:

<....> End-to-End Communication
<----> Hop-by-Hop Communication

Figure 2: Example Instantiation of involved Entities

We can distinguish between three types of adversaries:

Eavesdropper: RFC 4949 describes the act of 'eavesdropping' as

> "Passive wiretapping done secretly, i.e., without the knowledge of the originator or the intended recipients of the communication."

Eavesdropping is often considered by IETF protocols in the context of a security analysis to deal with a range of attacks by offering confidentiality protection.

RFC 3552 provides guidance on how to write security considerations for IETF documents and already demands the confidentiality security services to be considered. While IETF protocols offer guidance on how to secure communication against eavesdroppers deployments sometimes choose not to enable its usage.

Middleman: Many protocols developed today show a more complex communication pattern than just client and server communication, as motivated in Figure 2. Store-and-forward protocols are examples where entities participate in the message delivery even though they are not the final recipients. Often, these intermediaries only need to see a small amount of information necessary for message routing and security and/or protocol mechanisms should ensure that end-to-end information is made inaccessible for these entities. Unfortunately, the difficulty to deploy end-to-end security proceduces, the additional messaging, the computational overhead, and other business / legal requirements often slow down or prevent the deployment of these end-to-end security mechanisms giving these intermediaries more exposure to communication patters and communication payloads than necessary.

Recipient: Although it is not intuitive to treat the recipient as an adversary since the entire purpose of the communication interaction is to provide information to it. However, the degree of familiarity and the type of information that needs to be shared with such an entity may vary from context to context and also between application scenarios. Often enough, the sender has no strong familiarity with the other communication endpoint. While it seems to be advisable to utilize access control before disclosing information with such an entity reality in Internet communication practical reality is different. As such, a sender may still want to limit the amount of information disclosed to the recipient and some mutual understanding of what the purpose of the collection is, how long the personal data needs to be stored, and how processing takes place. Additionally, an important part of privacy protection is for the recipient to offer privacy notices on the usage of the collected personal data, to offer choices, and to obtain consent from the data subject.

4. Guidelines

A pre-condition for reasoning about the impact of a protocol or an architecture is to look at the high level protocol model, as described in [RFC4101]. This step helps to identify actors and their relationship. The protocol specification (or the set of specifications then allows a deep dive into the data that is exchanged.

The answers to these questions provide insight into the potential privacy impact:

1. What entities collect and use data?

   1.a: How many entities collect and use data?

   Note that this question aims to raise the question of what is possible for various entities to inspect (or potentially modify). In architectures with intermediaries, the question can be stated as "What data is exposed to intermediaries that they do not need to know to do their job?".

   1.b: For each entity, what type of entity is it?

   + The first-party site or application

   + Other sites or applications whose data collection and use is in some way controlled by the first party

   + Third parties that may use the data they collect for other purposes

2. For each entity, think about the relationship between the entity and the user.

   2.a: What is the user's familiarity or degree of relationship with the entity in other contexts?

   2.b: What is the user's reasonable expectation of the entity's involvement?

3. What data about the user is likely needed to be collected?

4. What is the identification level of the data? (identified, pseudonymous, anonymous, see [I-D.hansen-privacy-terminology])

The questions in this section are based on the CDT published paper

"Threshold Analysis for Online Advertising Practices" [CDT].

5. Example

This section allows us to illustrate how privacy was deal within certain IETF protocols. We will start the description with AAA for network access and expand it to other protocols in a future version of this draft.

5.1. Presence

A presence service, as defined in the abstract in RFC 2778 [RFC2778], allows users of a communications service to monitor one another's availability and disposition in order to make de- cisions about communicating. Presence information is highly dynamic, and generally characterizes whether a user is online or offline, busy or idle, away from communications devices or nearby, and the like. Necessarily, this information has certain privacy implications, and from the start the IETF approached this work with the aim to provide users with the controls to determine how their presence information would be shared. The Common Profile for Presence (CPP) [RFC3859] defines a set of logical operations for delivery of presence information. This abstract model is applicable to multiple presence systems. The SIP-based SIMPLE presence system [RFC3261] uses CPP as its baseline architecture, and the presence operations in the Extensible Messaging and Presence Protocol (XMPP) have also been mapped to CPP [RFC3922].

SIMPLE [RFC3261], the application of the Session Initiation Protocol (SIP) to instant messaging and presence, has native support for subscriptions and notifications (with its event framework [RFC3265]) and has added an event package [RFC3856] for pres- ence in order to satisfy the requirements of CPP. Other event packages were defined later to allow additional information to be exchanged. With the help of the PUBLISH method [RFC3903]. clients are able to install presence information on a server, so that the server can apply access-control policies before sharing presence information with other entities. The integration of an explicit authorization mechanism into the presence architecture has been a major improvement in terms of involving the end users in the decision making pro- cess before sharing information. Nearly all presence systems deployed today provide such a mechanism, typically through a reciprocal authorization system by which a pair of users, when they agree to be "buddies," consent to divulge their presence information to one another.

One important extension for presence was to enable the support for location sharing. With the desire to standardize protocols for systems sharing geolocation IETF work was started in the GEOPRIV working group. During the initial requirements and privacy threat analysis in the process of chartering the working group, it became

clear that the system would an underlying communication mechanism supporting user consent to share location information. The resemblance of these requirements to the presence framework was quickly recognized, and this design decision was documented in RFC 4079 [RFC4079].

While presence systems exerted influence on location pri- vacy, the location privacy work also influenced ongoing IETF work on presence by triggering the standardization of a general access control policy language called the Common Policy (defined in RFC 4745 [RFC4745]) framework. This language allows one to express ways to control the distribution of information as simple conditions, actions, and transformations rules expressed in an XML format. Common Policy itself is an abstract format which needs to be instantiated: two examples can be found with the Presence Authorization Rules [RFC5025] and the Geolocation Policy [I-D.ietf-geopriv-policy]. The former provides additional expressiveness for presence based systems, while the latter defines syntax and semantic for location based conditions and transformations.

As a component of the prior work on the presence architecture, a format for presence information, called Presence Information Data Format (PIDF), had been developed. For the purposes of conveying location information an extension was developed, the PIDF Location Object (PIDF-LO). With the aim to meet the privacy requirements defined in RFC 2779 [RFC2779] a set of usage indications (such as whether retransmission is allowed or when the retention period expires) in the form of the following policies have been added that always travel with location information itself. We believe that the standardization of these meta-rules that travel with location information has been a unique contribution to privacy on the Internet, recognizing the need for users to express their preferences when information travels through the Internet, from website to website. This approach very much follows the spirit of Creative Commons [CC], namely the usage of a limited number of conditions (such as 'Share Alike' [CC-SA]). Unlike Creative Commons, the GEOPRIV working group did not, however, initiate work to produce legal language nor to de- sign graphical icons since this would fall outside the scope of the IETF. In particular, the GEOPRIV rules state a preference on the retention and retransmission of location information; while GEOPRIV cannot force any entity receiving a PIDF-LO object to abide by those preferences, if users lack the ability to express them at all, we can guarantee their preferences will not be honored.

While these retention and retransmission meta-data elements could have been devised to accompany information elements in other IETF protocols, the decision was made to introduce these elements for

geolocation initially because of the sensitivity of location information.

The GEOPRIV working group had decided to clarify the architecture to make it more accessible to those outside the IETF, and also provides a more generic description applicable beyond the context of presence. [I-D.ietf-geopriv-arch] shows the work-in-progress writeup.

5.2. AAA for Network Access

On a high-level, AAA for network access uses the communication model shown in Figure 3. When an end host requests access to the network it has to interact with a Network Access Server (NAS) using some front-end protocol (often at the link layer, such as IEEE 802.1X). When asked by the NAS, the end host presents a Network Access Identifier (NAI), an email alike identifier that consists of a username and a domain part. This NAI is then used to discover the AAA server authorized for the users' domain and an initial access request is forwarded to it. To deal with various security, accounting and fraud prevention aspects an end-to-end authentication procedure, run between the end host (the peer) and a separate component within the AAA server (the server) is executed using the Extensible Authentication Protocol (EAP). After a successful authentication protocol exchange the user may get authorized to access the network and keying material is provided to the NAS to enable link layer security over the air interface.

From a privacy point of view, the entities participating in this eco-system are the user, an end host, the NAS, a range of different intermediaries, and the AAA server. The user will most likely have some form of contractual relationship with the entity operating the AAA server since credential provisioning had to happen someone but, in certain deployments like coffee shops, this is not guaranteed. In many deployment during this initial registration process the subscriber is provided with credentials after showing some form of identification information (e.g. a passport) and consequently the NAI together with credentials can be used to linked to a specific subscriber, often a single person.

The username part of the NAI is data provided by the end host provides during network access authentication that intermediaries do not need to fulfill their role in AAA message routing. Hiding the user's identity is, as discussed in RFC 4282 [RFC4282], possible only when NAIs are used together with a separate authentication method that can transfer the username in a secure manner. Such EAP methods have been designed and requirements for offering such functionality have have become recommended design criteria, see [RFC4017].

More than just identity information is exchanged during the network access authentication is exchanged. The NAS provides information about the user's point of attachment towards the AAA server and the AAA server in response provides data related to the authorization decision back. While the need to exchange data is motivated by the service usage itself there are still a number of questions that could be asked, such as

- What mechanisms can be utilized to offer users ways to authorize sharing of information (considering that the ability for protocol interaction is limited without sucessful network access connectivity)?

- What are the best current practices for a privacy-sensitive operation of intermediaries? Since end hosts are not interacting with intermediaries explicitly and users have no relationship with those who operate them it is quite likely their practices are less widely known.

- Are there alternative approaches to trust establishment between the NAS and the AAA server so that the involvement of intermediaries can be limited or avoided?

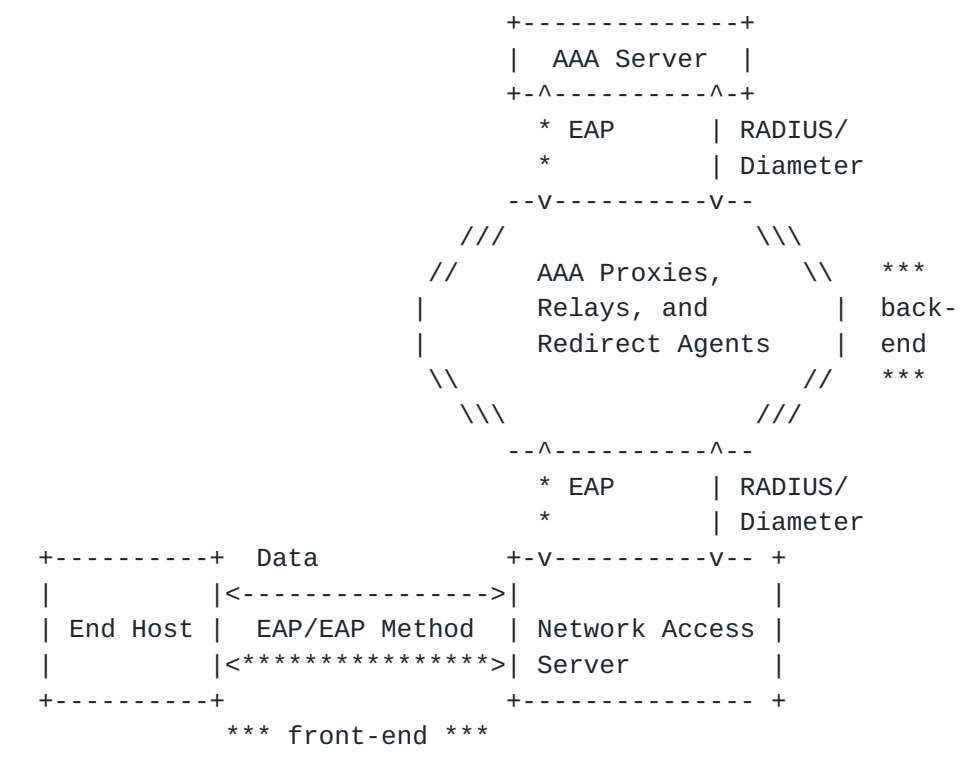


Figure 3: Network Access Authentication Architecture

6. Security Considerations

This document describes aspects a protocol designer would considered in the area of privacy in addition to the regular security analysis.

7. IANA Considerations

This document does not require actions by IANA.

8. Acknowledgements

We would like to thank the participants for the feedback they provided during the December 2010 Internet Privcy workshop co-organized by MIT, ISOC, W3C and IAB.

9. References

9.1. Normative References

[I-D.hansen-privacy-terminology]
Pfitzmann, A., Hansen, M., and H. Tschofenig, "Terminology for Talking about Privacy by Data Minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management", draft-hansen-privacy-terminology-01 (work in progress), August 2010.

[OECD] Organization for Economic Co-operation and Development, "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data", available at (September 2010) , http://www.oecd.org/EN/document/0,,EN-document-0-nodirectorate-no-24-10255-0,00.html, 1980.

9.2. Informative References

[Altman] Altman, I., "The Environment and Social Behavior: Privacy, Personal Space, Territory, Crowding", Brooks/Cole , 1975.

[CC] "Creative Commons", June 2010.

[CC-SA] "Creative Commons - Licenses", June 2010.

[CDT] Center for Democracy & Technology, "Threshold Analysis for Online Advertising Practices", available at http://www.cdt.org/privacy/20090128threshold.pdf, Jan 2009.

[CTIA] CTIA, "Best Practices and Guidelines for Location-Based Services", , March 2010.

[DPD95] European Commission, "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data", Official Journal L 281 , 23/11/1995 P. 0031 - 0050, November 2005.

[EFF-Privacy]
Blumberg, A. and P. Eckersley, "On Locational Privacy, and How to Avoid Losing it Forever", August 2009.

[Granada] International Working Group on Data Protection in

Telecommunications, "The Granada Charter of Privacy in a Digital World, Granada (Spain)", April 2010.

[I-D.ietf-ecrit-framework]
Rosen, B., Schulzrinne, H., Polk, J., and A. Newton, "Framework for Emergency Calling using Internet Multimedia", draft-ietf-ecrit-framework-12 (work in progress), October 2010.

[I-D.ietf-geopriv-arch]
Barnes, R., Lepinski, M., Cooper, A., Morris, J., Tschofenig, H., and H. Schulzrinne, "An Architecture for Location and Location Privacy in Internet Applications", draft-ietf-geopriv-arch-03 (work in progress), October 2010.

[I-D.ietf-geopriv-policy]
Schulzrinne, H., Tschofenig, H., Morris, J., Cuellar, J., and J. Polk, "Geolocation Policy: A Document Format for Expressing Privacy Preferences for Location Information", draft-ietf-geopriv-policy-22 (work in progress), October 2010.

[I-D.morris-policy-cons]
Morris, J., Aboba, B., Peterson, J., and H. Tschofenig, "Public Policy Considerations for Internet Protocols", draft-morris-policy-cons-00 (work in progress), October 2010.

[Madrid] Data Protection Authorities and Privacy Regulators, "The Madrid Resolution, International Standards on the Protection of Personal Data and Privacy", Conference of Data Protection and Privacy Commissioners , 31st International Meeting, November 2009.

[RFC2778] Day, M., Rosenberg, J., and H. Sugano, "A Model for Presence and Instant Messaging", RFC 2778, February 2000.

[RFC2779] Day, M., Aggarwal, S., Mohr, G., and J. Vincent, "Instant Messaging / Presence Protocol Requirements", RFC 2779, February 2000.

[RFC2804] IAB and IESG, "IETF Policy on Wiretapping", RFC 2804, May 2000.

[RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261,

June 2002.

[RFC3265] Roach, A., "Session Initiation Protocol (SIP)-Specific Event Notification", RFC 3265, June 2002.

[RFC3856] Rosenberg, J., "A Presence Event Package for the Session Initiation Protocol (SIP)", RFC 3856, August 2004.

[RFC3859] Peterson, J., "Common Profile for Presence (CPP)", RFC 3859, August 2004.

[RFC3903] Niemi, A., "Session Initiation Protocol (SIP) Extension for Event State Publication", RFC 3903, October 2004.

[RFC3922] Saint-Andre, P., "Mapping the Extensible Messaging and Presence Protocol (XMPP) to Common Presence and Instant Messaging (CPIM)", RFC 3922, October 2004.

[RFC4017] Stanley, D., Walker, J., and B. Aboba, "Extensible Authentication Protocol (EAP) Method Requirements for Wireless LANs", RFC 4017, March 2005.

[RFC4079] Peterson, J., "A Presence Architecture for the Distribution of GEOPRIV Location Objects", RFC 4079, July 2005.

[RFC4101] Rescorla, E. and IAB, "Writing Protocol Models", RFC 4101, June 2005.

[RFC4282] Aboba, B., Beadles, M., Arkko, J., and P. Eronen, "The Network Access Identifier", RFC 4282, December 2005.

[RFC4745] Schulzrinne, H., Tschofenig, H., Morris, J., Cuellar, J., Polk, J., and J. Rosenberg, "Common Policy: A Document Format for Expressing Privacy Preferences", RFC 4745, February 2007.

[RFC4858] Levkowetz, H., Meyer, D., Eggert, L., and A. Mankin, "Document Shepherding from Working Group Last Call to Publication", RFC 4858, May 2007.

[RFC4962] Housley, R. and B. Aboba, "Guidance for Authentication, Authorization, and Accounting (AAA) Key Management", BCP 132, RFC 4962, July 2007.

[RFC5025] Rosenberg, J., "Presence Authorization Rules", RFC 5025, December 2007.

[RFC5598] Crocker, D., "Internet Mail Architecture", RFC 5598, July 2009.

[SP800-122] McCallister, E., Grance, T., and K. Scarfone, "Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)", NIST Special Publication (SP) , 800-122, April 2010.

[SP800-130] Barker, E., Branstad, D., Chokhani, S., and M. Smid, "DRAFT: A Framework for Designing Cryptographic Key Management Systems", NIST Special Publication (SP) , 800-130, June 2010.

[Tussle] Clark, D., Wroslawski, J., Sollins, K., and R. Braden, "Tussle in Cyberspace: Defining Tomorrow's Internet", In Proc. ACM SIGCOMM , http://www.acm.org/sigcomm/sigcomm2002/papers/tussle.html, 2002.

[Warren] Warren, D. and L. Brandeis, "The Right to Privacy", Harvard Law Rev. , vol. 45, 1890.

[Westin] Westin, A., "Privacy and Freedom", Atheneum, New York , 1967.

[browser-fingerprinting] Eckersley, P., "How Unique Is Your Browser?", Springer Lecture Notes in Computer Science , Privacy Enhancing Technologies Symposium (PETS 2010), 2010.

[limits] Cate, F., "The Limits of Notice and Choice", IEEE Computer Society , IEEE Security and Privacy, pg. 59-62, November 2005.

Appendix A. Historical Background

The "right to be let alone" is a phrase coined by Warren and Brandeis in their seminal Harvard Law Review article on privacy [Warren]. They were the first scholars to recognize that a right to privacy had evolved in the 19th century to embrace not only physical privacy but also a potential "injury of the feelings", which could, for example, result from the public disclosure of embarrassing private facts.

In 1967 Westin [Westin] described privacy as a "personal adjustment process" in which individuals balance "the desire for privacy with the desire for disclosure and communication" in the context of social norms and their environment. Privacy thus requires that an individual has a means to exercise selective control of access to the self and is aware of the potential consequences of exercising that control [Altman].

Efforts to define and analyze the privacy concept evolved considerably in the 20th century. In 1975, Altman conceptualized privacy as a "boundary regulation process whereby people optimize their accessibility along a spectrum of 'openness' and 'closedness' depending on context" [Altman]. "Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others. Viewed in terms of the relation of the individual to social participation, privacy is the voluntary and temporary withdrawal of a person from the general society through physical or psychological means, either in a state of solitude or small-group intimacy or, when among larger groups, in a condition of anonymity or reserve." [Westin].

Note: Altman and Westin were referring to non-electronic environments, where privacy intrusion was typically based on fresh information, referring to one particular person only, and stemming from traceable human sources. The scope of possible privacy breaches was therefore rather limited. Today, details about an individual's activities are typically stored over a longer period of time, collected from many different sources, and information about almost every activity in life is available electronically.

In 1980, the Organization for Economic Co-operation and Development (OECD) published eight Guidelines on the Protection of Privacy and Trans-Border Flows of Personal Data [OECD], which are often referred to as Fair Information Practices (FIPs). Fair information practices include the following principles:

Notice and Consent: Before the collection of data, the data subject should be provided: notice of what information is being collected and for what purpose and an opportunity to choose whether to accept the data collection and use. In Europe, data collection cannot proceed unless data subject has unambiguously given his consent (with exceptions).

Collection Limitation: Data should be collected for specified, explicit and legitimate purposes. The data collected should be adequate, relevant and not excessive in relation to the purposes for which they are collected.

Use/Disclosure Limitation: Data should be used only for the purpose for which it was collected and should not be used or disclosed in any way incompatible with those purposes.

Retention Limitation: Data should be kept in a form that permits identification of the data subject no longer than is necessary for the purposes for which the data were collected.

Accuracy: The party collecting and storing data is obligated to ensure its accuracy and, where necessary, keep it up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete are corrected or deleted.

Access: A data subject should have access to data about himself, in order to verify its accuracy and to determine how it is being used.

Security: Those holding data about others must take steps to protect its confidentiality.

The OECD guidelines and also more recent publications, like the Madrid resolution [Madrid] or the Granada Charter of Privacy in a Digital World [Granada], provide a useful understanding of how to provide privacy protection but these guidelines quite naturally stay on a higher level. They are idealistic principles. As such, they do not aim to evaluate the tradeoffs in addressing privacy protection in the different stages of the development process, as illustrated in Figure 1.

US regulatory and self-regulatory efforts supported by the Federal

Trade Commission (FTC) have focused on a subset of these principles, namely to notice, choice, access, and security rather than minimizing data collection or use limitation. Hence, they are sometimes labeled as the "notice and choice" approach to privacy. From a practical point of view it became evident that companies are reluctant to stop collecting and using data but individuals expect to remain in control about its usage. Today, the effectiveness to deal with privacy violations using the "notice and choice" approach is heavily criticized [limits].

Among these considers (although often implicit) are assumptions on how information is exchanged between different parties and for certain protocols this information may help to identify entities, and potentially humans behind them. Without doubt the information exchanged is not always equal. The terms 'personal data' [DPD95] and Personally Identifiable Information (PII) [SP800-122] have become common language in the vocabulary of privacy experts. It seems therefore understandable that regulators around the globe have focused on the type of data being exchanged and have provided laws according to the level of sensitivity. Medical data is treated differently in many juristictions than blog comments. For an initial investigation it is intuitive and helpful to determine whether specific protocol or application may be privacy sensitive. The ever increasing ability for parties on the Internet to collect, aggregate, and to reason about information collected from a wide range of sources requires to apply further thinking about potential other privacy sensitive items. The recent example of browser fingerprinting [browser-fingerprinting] shows that tracking can happen in surprising ways.

The following list contains examples of information that may be considered personal data:

- Name
- Address information
- Phone numbers, email addresses, SIP/XMPP URIs, other identifiers
- IP and MAC addresses or other host-specific persistent identifiers that consistently links to a particular person or small, well-defined group of people
- Information identifying personally owned property, such as vehicle registration number

Searching only for those example as an indication for the need of privacy is, however, insufficient given that the list above is

constantly growing and depends very much on the context. An information element may not be sensitive in one context but considered very sensitive in another. In aggregation possibilities have also caused the list of personal data to grow.

Authors' Addresses

Bernard Aboba
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
US

Email: bernarda@microsoft.com

John B. Morris, Jr.
Center for Democracy and Technology
1634 I Street NW, Suite 1100
Washington, DC 20006
USA

Email: jmorris@cdt.org
URI: http://www.cdt.org

Jon Peterson
NeuStar, Inc.
1800 Sutter St Suite 570
Concord, CA 94520
US

Email: jon.peterson@neustar.biz

Hannes Tschofenig
Nokia Siemens Networks
Linnoitustie 6
Espoo 02600
Finland

Phone: +358 (50) 4871445
Email: Hannes.Tschofenig@gmx.net
URI: http://www.tschofenig.priv.at