Privacy Considerations for Internet Protocols

Bringing Privacy to the Internet Engineering Task Force (IETF)

Hannes Tschofenig[1]

[1]Nokia Siemens Networks

I. Learning from Security

”Privacy by Design” has become a favorable term among regulators, researchers, and engineers. Those who have followed the work on security will find the idea of addressing privacy early in the design process as intuitive rather than incorporating privacy functionality as an afterthought. Without doubt the security community has gone a long way by agreeing on terminology, defining threat models, and identifying basic security properties. The Request for Comments (RFCs) of the Internet Engineering Task Force (IETF) on ”Guidelines for Writing RFC Text on Security Considerations” (RFC 3552 [2]), the ”Internet Security Glossary” (RFC 4949 [3]), and ”Writing Protocol Models” (RFC 4101 [4]) are among the documents often cited when talking about addressing security in the design of protocols and architectures. These documents have given engineers a lot of insight but the organizational structure that was setup to support security reviews has proven essential for improving security. As an example the author will illustrate how the process was executed in the IETF, which is an open standards developing organization.

All RFCs are required to have a Security Considerations section, as stated in RFC 1543 [5] (and updated by RFC 2223 [6]). At the beginning the quality of the writeup in those security considerations sections was relatively weak, as confirmed in [7]. But, neither RFC provides much guidance: ”All RFCs must contain a section near the end of the document that discusses the security considerations of the protocol or procedures that are the main topic of the RFC.” With the introduction of this mandatory security analysis a separate review group, called ”Security Area Directorate” [8], was introduced. This group consists of the working group chairs of the security area and selected individuals chosen for their technical knowledge in security. They review every IETF document primarily to help the area directors improve their efficiency in the document approval process. In practice, these review comments are taken serious by document authors and delays in publication have not been uncommon when security vulnerabilities have been discovered. This group also reviews documents early in the process when problems are anticipated or working group chairs have asked for a security advisor. To facilitate the exchange of information among the participants a face-to-face meeting takes place at every

[*] This paper is a submission to the 49th meeting of the International Working Group on Data Protection in Telecommunications, 4 - 5 April 2011, Montreal, Canada. The content represents the views of the author in his role as individual contributor to the Internet standards process. They do not reflect the consensus output of any IETF working group or the Internet Architecture Board (IAB). Referenced documents may, however, reflect IETF consensus depending on the type of the document. More details about the IETF standards process is described in [1]. Please address correspondence to Hannes.Tschofenig@nsn.com.

IETF meeting and presentations are given to the broader IETF security community on challenging topics at the Security Area Directorate Open Meeting.

Finding security experts who volunteer to spend a significant amount of time to review long and detailed technical specifications, and to discuss their findings with the authors and the respective working group is difficult. For many experts this activity is not as rewarding as publishing papers and speaking at conferences.

On top of the above-mentioned process the IETF EDU team [9] organizes security education training sessions at the beginning of IETF meetings. Based on the limited time these tutorials are considered food for thought for newcomers rather than an attempt to train security protocol experts.

II. From Security to Privacy

While a number of standardized security technologies have found widespread deployment the same observation cannot be observed in the area of privacy (yet). There is a long list of privacy enabling mechanisms that have been developed by the standardization and the research community but how many of them have found their way into deployments? Clearly, something has gone wrong. As a consequence, the author believes that there is not a lot of experience in engineering privacy into technical systems (within the standards community).

To start somewhere the author believes it is important to build on top of an already successful model[1] and our experience with security is what seems to be a sufficiently close match. The author is convinced that privacy has to be introduced in standards development organizations in the same way as it has been done with security. This has lead us to write a document about ”Privacy Considerations for Internet Protocols” [10] to make protocol designers aware of privacy-related design choices and to offer guidance for writing text for those IETF documents. This document aims to serve the same function as RFC 3552 [2]. For privacy terminology [11] was written to provide guidance. In response to the IAB Internet Privacy workshop [12] the security area directors have created a privacy directorate [13], following the style and purpose of the previously described security directorate.

Developing complex distributed systems takes a long time and typically involves multiple organizations. There is no single approach how successful technology gets developed but RFC 5218 [14] discusses what criteria are successful protocols have. At least, one has to differentiate the protocol and architecture specification from the actual deployment. Typically, there are gaps between the two; not only because the specification often leaves freedom in how to combine different components. Similarly, to incorporating security during the design there are a range of decisions that need to be made during the research, standardization, implementation and deployment phase. Below, two examples to illustrate this important point are provided below. First, we take an example from the security space followed by an example from the privacy area.

Email Security

[1] There is also some disagreement in the security community whether the work on security has indeed been so successful given the long list of still unresolved security challenges on the Internet. This aspect is, however, not investigated in this article.

> Email is a fairly old technology and a number of security mechanisms have been defined over time addressing the evolving Internet threat model. For example, RFC 2595 [15] specifies how Transport Layer Security (TLS) is used with Internet Message Access Protocol (IMAP) [16] and the Post Office Protocol (POP) [17], two email protocols often used for retrieving and sending emails. RFC 2595 was published in 1999 and still today a number of email providers do not offer protection of the email client-server interaction via TLS or do not enable it per default. With the introduction of email access via the Web browser this situation has not improved either. Instead, many users still send and receive emails over an unsecured communication channel allowing adversaries (for example on a public Wifi hotspots) to eavesdrop on the communication.
>
> **Open Web Authentication (OAuth)**
>
> The IETF OAuth working group [18] develops protocols for secure data sharing on the Web and allows one website to retrieve data stored at another website on behalf of a user. The classical example is the photo printing site that wants to access the private pictures of a user stored at some other picture sharing website (with the user's consent). While the OAuth specification [19] defines a protocol exchange for requesting and obtaining authorization tokens it does not make attempts to understand the semantic of the shared data. As such, it does not make a difference for OAuth whether travel plans, pictures, or medical records are being shared. It is accepted protocol engineering practice to design for extensibility, see also Section 2.2.1 of RFC 5218 [14] which observes that extensibility is one important criteria for successful protocol adoption. From a protocol point of view it is only exchanging data and the data could be anything. Additionally, the way how a website deploys OAuth, implements the access control model, how the user interface is designed, and what other operational procedures that site adheres to, is largely beyond the scope of protocol and architecture standardization. End users and applications using OAuth are typically only exposed to a very small slice, often only those that have some impact to the user-visible components. Quite naturally, design shortcomings in the user-visible aspects quickly receive a lot of media attention. An example of such recent press attention was causes by a blog post entitled as ”OAuth Will Murder Your Children” [20].

These two examples aim to illustrate that the protocol and architectural design in a standards developing organization is necessary, but clearly not sufficient, to ensure a successful privacy-aware system for even the most basic and obvious privacy intrusions. Instead, the costs and the incentives need to be aligned properly to consider privacy throughout the entire design process.

Privacy is only one of many design criteria and competes with many other, such as features functionality, performance, flexibility and extensibility, security, manageability and configurability, usability, or just basic design aspects just as layering, NAT and firewall traversal, naming and addressing, service discovery, congestion control, or internationalization support. Design philosophies also play an important role since the engineering process is a form of art and not just a purely mechanical process. For example, a few years ago the research and the engineering communities were excited about peer-to-peer networking and utilizing the distributed nature of communication, for example using Distributed Hash Tables (DHTs) was commonly found in architectural designs. Distributing information and to not maintain them centrally is, in privacy terms, often considered desirable. A few years later the design spirit in the industry has, however, changed again and the main theme of today is cloud computing where storing data on the server side (rather than distributing them among end hosts) is not undesirable anymore. Needless to say that these trends in the industry are supported by the cost and price shifts in the computing infrastructure, for example cost of storage and cost of transmission. In any case, these patterns are strongly reflected in application design and they impact the privacy in subtle ways.

III. From Data Minimalization to Threshold Analysis

What guidelines can be given to a protocol designer with regard to privacy? [10] tries to provide an answer to this question. While it initially seemed to be obvious to make use of the "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data" [21], and particularly the collection limitation principle (also known as data minimalization principle).

Unfortunately, this approach has turned out to be challenging. The main reason is in the desire in the standardization community to identify generic building blocks that can be used throughout a number of different architectural environments and application domains. This is in strong contrast to what is being exercised in research papers where complete systems are designed (often by a small group driven by the same incentives). In a position paper to the 'Real-Time Communication in the Browser' workshop [22] we describe our observations that the role of standardization is changing dramatically. The increased speed of innovation demands more flexibility in the standards process. Organizations who focus on a 'standardize the complete system' will struggle in the market place.

> To illustrate our point we would like to pick a random example. In [23] A. Blumberg and P. Eckersley point to a number of examples of privacy respective applications, including a road pricing application. VPriv [24] is one such listed research idea for road pricing that aims to offer better privacy properties. Researchers are free to define their own design constraints and requirements. Therefore problems with conflicting interests in design goals often do not surface since they can be 'defined away' fairly easily. A standardization organization on the other hand would more likely focus on a mechanism to convey location information as a building block, such as provided by [25], rather than standardizing a road pricing system altogether. Even such a simple technical task, such as conveying location from one communication end point to the other, can very easily turn into a several year effort in a standards developing organization. Coming to an agreement on the privacy characteristics can even be more difficult.

As such, the purpose of the developed building blocks is to typically to allow the exchange of information for a large number of foreseen usage scenarios, but also purposes unknown at the time of the design. Designing for the unknown use cases is indeed an important part of the overall design. For example, HTTP was developed to transfer arbitrary data and is widely used today by application developers on the Internet for radically different purposes. HTTP is an example of a wildly successful protocol.

Instead of designing for data minimalization and purpose limitation we argue for a different approach that is more intuitive for our engineers. In [10] we attempted to generalize the work done by the Center for Democracy and Technology (CDT) on ”Threshold Analysis for Online Advertising Practices” [26], [27].

In a nutshell, the currently proposed approach requires engineers to describe the privacy properties of their protocols and architectures following a few basic questions:

1) What entities collect and use data?
    a) How many entities collect and use data?
    b) For each entity, what type of entity is it?
2) For each entity, think about the relationship between the entity and the user.
    a) What is the user’s familiarity or degree of relationship with the entity in other contexts?
    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?

More details can be found in [10]. This template had been applied in recent IETF work [28].

Our expectation is that the thought-process of analyzing the privacy properties of the designed system will allow working group participants to investigate various solution approaches to make more informed decisions and to hopefully favor those solutions that provide better privacy characteristics. We also believe that this analysis will help to discover stupid privacy mistakes (such as the bug that was introduced with the original design of the IPv6 stateless address autoconfiguration mechanism where the MAC address was incorporated into the interface identifier part of the IPv6 address).

IV. Conclusions

The concept of ’Privacy by Design’ has gotten a lot of attention of the past few years and within the IETF we have tried to investigate how we can consider privacy in the design of protocols and architectures in a more systematic way. As argued in our position paper [29] to the W3C Workshop on Privacy for Advanced Web APIs [30] the IETF does consider privacy to a certain extend already in their protocols and architectural designs despite the lack of detailed guidelines. Nevertheless, there is room for improvement. We have started to shed more light on privacy in the IETF already by organizing a privacy workshop [12] to solicit input from the technically minded privacy community, to create an IETF privacy directorate, and to start the work on a number of documents to offer more guidance to engineers. More awareness and education activities will follow with the upcoming IETF meeting in Prague in March 2011.

Based on our experience in the standardization environment we do, however, face certain challenges that researchers are typically not constrained by. These challenges include the desire to design versatile building blocks rather than complete systems, the conflicting goals of different design criteria, the length of the standardization process, and the need to follow the consensus process.

The introduction of privacy into the design process will take time. To speed up the process we are aiming to follow the path taken by security. Since protocol design is complex and requires a lot of expertise in different areas we are spending extended focus on conveying a simple message, as we tried to highlight with the work on the ”Privacy Considerations for Internet Protocols” document [10].

REFERENCES

[1] S. Bradner, “ The Internet Standards Process – Revision 3,” Oct. 1996, RFC 2026, Request For Comments.

[2] E. Rescorla and B. Korver, “Guidelines for Writing RFC Text on Security Considerations,” Jul. 2003, RFC 3552, Request For Comments.

[3] R. Shirey, “Internet Security Glossary, Version 2,” Aug. 2007, RFC 4949, Request For Comments.

[4] E. Rescorla, “Writing Protocol Models,” Jun. 2005, RFC 4101, Request For Comments.

[5] J. Postel, “Instructions to RFC Authors,” Oct. 1993, RFC 1543, Request For Comments.

[6] J. Postel and J. Reynolds, “Instructions to RFC Authors,” Oct. 1997, RFC 2223, Request For Comments.

[7] A. Rabkin, N. Doty, and D. K. Mulligan, “Facilitate, don’t mandate,” Dec. 2010, position Paper for the IAB/W3C/MIT/ISOC Internet Privacy Workshop, Boston, Dec. 2010, http://www.iab.org/about/workshops/privacy/papers/nick_doty.pdf.

[8] IETF, “IETF Security Directorate Wiki,” Feb. 2011, http://bit.ly/hZEylU.

[9] ——, “Education (EDU) Team Wiki,” Feb. 2011, http://wiki.tools.ietf.org/group/edu/.

[10] B. Aboba, J. Morris, J. Peterson, and H. Tschofenig, “Privacy Considerations for Internet Protocols,” Nov. 2010, IETF draft (work in progress), draft-morris-privacy-considerations-02.txt.

[11] A. Pfitzmann, M. Hansen, and H. Tschofenig, “Terminology for Talking about Privacy by Data Minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management,” Aug. 2010, IETF draft (work in progress), draft-hansen-privacy-terminology-01.txt.

[12] IAB, W3C, ISOC, and MIT, “Internet Privacy Workshop: How can Technology help to improve Privacy on the Internet?” Dec. 2010, http://www.iab.org/about/workshops/privacy/.

[13] IETF, “IETF Announcement of the Privacy Directorate,” Feb. 2011, http://www.ietf.org/mail-archive/web/ietf-announce/current/msg08294.html.

[14] D. Thaler and B. Aboba, “What Makes for a Successful Protocol?” Jul. 2008, RFC 5218, Request For Comments.

[15] C. Newman, “Using TLS with IMAP, POP3 and ACAP,” Jun. 1999, RFC 2595, Request For Comments.

[16] M. Crispin, “INTERNET MESSAGE ACCESS PROTOCOL - VERSION 4rev1,” Mar. 2003, RFC 3501, Request For Comments.

[17] J. Myers and M. Rose, “Post Office Protocol - Version 3,” May 1996, RFC 1939, Request For Comments.

[18] IETF, “Open Authentication Protocol (oauth) Working Group Charter,” Feb. 2011, http://datatracker.ietf.org/wg/oauth/charter/.

[19] E. Hammer-Lahav, D. Recordon, and D. Hardt, “The OAuth 2.0 Authorization Protoco,” Jan. 2011, IETF draft (work in progress), draft-ietf-oauth-v2-12.txt.

[20] Z. Holman, “OAuth Will Murder Your Children,” Jan. 2011, http://zachholman.com/2011/01/oauth_will_murder_your_children.

[21] “OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data,” 1980, http://www.oecd.org/EN/document/0,,EN-document-0-nodirectorate-no-24-10255-0,00.html.

[22] B. Aboba, J. Peterson, H. Schulzrinne, and H. Tschofenig, “The Future of Web Applications or How to Move into the Post Standardization Area,” Oct. 2010, position Paper for the ’Real-Time Communication in the Browser’ workshop, Mountain View,.

[23] A. Blumberg and P. Eckersley, “On Locational Privacy, and How to Avoid Losing it Forever,” Aug. 2009, http://www.eff.org/wp/locational-privacy.

[24] R. A. Popa, H. Balakrishnan, and A. J. Blumberg, “VPriv: Protecting privacy in location-based vehicular services,” in Proceedings of the 18th Usenix Security Symposium, August 2009. [Online]. Available: http://www.usenix.org/events/sec/tech/full_papers/popa.pdf

[25] J. Polk, B. Rosen, and J. Peterson, “Location Conveyance for the Session Initiation Protocol,” Oct. 2010, IETF draft (work in progress), draft-ietf-sipcore-location-conveyance-04.txt.

[26] CDT, “Threshold Analysis for Online Advertising Practices,” Jan. 2009, http://www.cdt.org/privacy/20090128threshold.pdf,.

[27] A. Cooper and J. Morris, “Thoughts on Adding ”Privacy Considerations” to Internet Drafts,” Dec. 2010, position Paper for the IAB/W3C/MIT/ISOC Internet Privacy Workshop, Boston, Dec. 2010, http://www.iab.org/about/workshops/privacy/papers/alissa_cooper.pdf.

[28] J. Howlett, S. Hartman, H. Tschofenig, and E. Lear, “Application Bridging for Federated Access Beyond Web (ABFAB) Architecture,” Mar. 2011, IETF draft (work in progress), draft-lear-abfab-arch-02.txt.

[29] J. Peterson, H. Tschofenig, B. Aboba, and K. Sollins, “The Role of the Internet Engineering Task Force (IETF) in Improving Privacy on the Internet,” Jul. 2010, position Paper for the W3C Workshop on Privacy for Advanced Web APIs, Jul. 2010, http://www.w3.org/2010/api-privacy-ws/papers/privacy-ws-32.pdf.

[30] W3C, “W3C Workshop on Privacy for Advanced Web APIs,” Jul. 2010, http://www.w3.org/2010/api-privacy-ws/.