Challenges for the Location-Aware Web

Alissa Cooper
Center for Democracy and Technology
acooper@cdt.org

Deirdre K. Mulligan
School of Information
UC Berkeley
dkm@ischool.berkeley.edu

Henning Schulzrinne
Department of Computer Science
Columbia University
hgs@cs.columbia.edu

Erik Wilde
School of Information
UC Berkeley
dret@berkeley.edu

ABSTRACT

The Web is on its way to becoming a location-aware information system. This transition causes some technical and policy challenges in terms of both design and coordination with existing approaches in this area. In this paper we propose that managing the transition to location-awareness (and some other aspects) requires a more strategic approach than has been taken thus far.

1. INTRODUCTION

This paper looks at the current state of affairs regarding location-enabling the Web. While location-based applications have become almost ubiquitous, the Web itself still is location-unaware. This deficiency has been pointed out repeatedly [16], and there still is no stable Web-level mechanism for location-oriented information. However, based on the immense popularity of web-enabled mobile devices, as a first development in that area an API for client-side location information access [24] is now under development by a *World Wide Web Consortium (W3C)* working group. Separate work is taking place under the purview of an IETF working group to standardize a geolocation URI format [19]. This latter working group, which has historically been oriented more towards real-time Internet applications (presence and VoIP) than Web applications, has been and continues to develop an architecture for location and location privacy in Internet applications [2, 8, 10].

This paper is an attempt to bring together the various communities involved in developing location-oriented specifications for the Web or for the Internet. The advent of mobile devices and a multitude of Web-based technologies and applications makes this a critical and historical period in the advancement of the Web towards a location-aware system. We argue that this is an area of high importance that requires strategic management and vision by the W3C. Location support is too critical of a concept to the future of the Web for it to be developed in a vacuum or tacked on as an afterthought to existing Web architectures. While the IETF's focus is not strictly on Web technologies, it has successfully taken on application-layer work in other areas, and the overlap of interest in the area of geolocation and location-oriented functionality is big enough to warrant a strategic partnership. Such a development would not only help to better align two hugely important Internet-based infrastructures, but it would also foster cooperation and an expanded understanding of the application area and possible use cases.

2. LOCATION AND NETWORKS

For a long time, the Internet almost by definition "ignored" location, as did the Web. One of the nice aspects about the Web was exactly the fact that it seamlessly connected clients and servers, and with the advent of social media, users and users, in a way that was completely independent of physical location. This was, from the very outset, a distinct departure from the other worldwide communications system, the *Public Switched Telephone Network (PSTN)*. The PSTN developed nationally and never grew past the stage of nationally operating entities with interconnection agreements, largely because of economic constraints (national operators with nationally limited customer bases).

The recent advent of the *Mobile Web*, which is a somewhat vague term but most often refers to the fact that modern mobile devices nowadays are full-fledged Web-capable clients, has spurred two interesting developments. On the one hand, the very distinct usage characteristics of mobile devices [18] have opened new opportunities for monetization, a development that often is summarized as *Location-Based services (LBS)* (a typical example is location-based advertising, for example listing nearby restaurants when users access the Web from their mobile device). On the other hand, most mobile devices nowadays also are telephones, which merges the previously distinct global communications networks. This causes some interesting side-effects, such as the question of how to uphold the traditional distinction between the telephone network and the Internet as the two networks merge. This merging is still in its infancy (demonstrated by how little the existing `tel` [25] and `sms` [28] URI schemes are used on the public Web, for example), and it will not happen very rapidly or without some setbacks, but it is likely safe to assume that at some point in time the existing separate network infrastructures will merge. This will profoundly impact fee structures for network access, and how exactly this restructuring will play out is almost impossible to predict.

Web Science Conf. 2010, April 26-27, 2010, Raleigh, NC, USA.
.

Another relevant new development is the *Web of Things* [12, 13]: the merging of pervasive and ubiquitous computing (embedded devices and sensor networks) with Web architectural principles. In some areas this is happening already, but in most cases these are still custom-built and fairly limited systems. In the next few years, however, information made available over the Web will likely originate in the Web of Things, and since so many everyday objects and environments today are sensor-equipped and network-enabled, this will open a fascinating foundation for new applications while also creating substantial challenges in controlling, regulating, and perhaps even legislating uses and misuses in this area.

Location will without a doubt continue to play an increasingly important role as all of these models converge. Unfortunately, the W3C does not seem to be approaching location on the Web strategically. For example, the issue of location privacy (how to deal with privacy issues around providing access to location information on personal devices) was deferred to the working group that is developing the first location-related Web standard, which is the *W3C Geolocation API* [24]. The working group mostly avoided the hard questions around how to best deal with this sensitive data [9], and it remains to be seen whether the rather lenient approach to privacy will be accepted by the W3C (the API is still in draft stage). The *W3C Device APIs and Policy Working Group* has been chartered "to create client-side APIs that enable the development of Web Applications and Web Widgets that interact with devices services such as Calendar, Contacts, Camera, etc. Additionally, the group will produce a framework for the expression of security policies that govern access to security-critical APIs (such as the APIs listed previously)." It remains to be seen how this framework will develop, and while the current approach is that geolocation is explicitly excluded from the group's charter (because there is a separate group that developed the geolocation draft), it is somewhat foreseeable that any framework that will be developed only makes sense if it is applied to the largest possible set of APIs.

Location information may very well also surface in places other than at the API level, such as in HTTP interactions or in URI schemes. Similar questions around privacy issues will arise in those scenarios, and it seems that the current approach of handling location privacy at the specification level for a single technology may not be a sustainable strategy. Current browser controls for privacy and security issues already are too complicated for most Web users, and adding specific mechanisms for each new technology that is brought to the Web might soon require more integrated approaches where users can express their preferences and settings and have them communicated to the user agent as well as to service providers.

3. STANDARDS AND PRIVACY

Attention to privacy in Web standards began in earnest in late 1994 when heated debate arose at the Internet Engineering Task Force over a proposal to standardize, within the Hypertext Transfer Protocol (HTTP), the ability for third parties to set "cookies" in users' browsers [17]. A nascent network advertising trade association objected to default settings that would impede its industry's business model by limiting third-party "cookies"; meanwhile, privacy advocates, the press and regulatory bodies objected to the invisibility of the data collection, the lack of user control, and the risks of data aggregation [17]. Beginning in 1996, the role of Web standards in enabling privacy protection became a central focus of research, debate and standards activity with the formation of the World Wide Web Consortium's *Platform for Privacy Preferences Project (P3P)* [6, 7, 17]. Shortly thereafter, the IETF's geographic location privacy (Geopriv) working group formed in recognition that "the representation and transmission of [location] information has significant privacy and security implications." In response to these concerns, Geopriv has "created a suite of protocols that allow such applications to represent and transmit such location objects and to allow users to express policies on how these representations are exposed and used."[1]

In a similar vein, efforts to consider the policy implications of rights expression languages analyzed their ability to support both rights and limitations provided for in copyright law [21]. At a higher level, several techniques for identifying and addressing the impact of technical designs on policy or social values have been offered with varying effects [3, 11, 20]. In the area of privacy in particular, researchers have directed attention to the ways in which technical design alters the norms of information flow [22], removes structural barriers that provide de facto privacy protection [26], increases visibility, transparency and exposure [4], introduces persistent identifiers, facilitates monitoring and tracking, and enables the collection and retention of information about individual users [20].

4. LOCATION INFORMATION PRIVACY

Information about location — both real-time location as well as permanent locations (such as home address) — garners special attention due to the consequences for both privacy and physical safety that may flow from its disclosure.

The heightened privacy and physical safety concerns generated by the collection, use and disclosure of location information are reflected in U.S. laws that create restrictive consent standards for its use and disclosure (47 USC §222), judicial concern about the standards governing law enforcement access to real-time and historical location information from telecommunications providers[2], limitations on the disclosure of department of motor vehicle records [27] and home addresses in public records [1], and requirements to offer caller-id suppression services.

At the level of norms, the heightened sensitivity with which individuals regard location information is reflected in social practices as well as empirical data. Parents and educators routinely remind children not to give out their home addresses or their physical locations. This advice is also strongly reflected in educational efforts aimed at digital youth, where a primary rule of the road is to never give out address or location information to strangers or post it publicly. It is also reflected in a trend to remove address information from phone directories, limit the availability of school contact lists, and other community efforts to protect the privacy of home addresses. Survey data confirms this evident sensitivity of location information. For example, a recent representative survey of Californians found strong

[1] `http://www.ietf.org/dyn/wg/charter/geopriv-charter.html`

[2] `http://www.eff.org/related/3494/pressrelease`

support for judicial intervention and due process before law enforcement are given access to historical location data (73% supporting a law requiring "police to convince a judge that a crime has been committed before obtaining location information from the cell phone company" [15]). Qualitative work exploring the use of location-sharing platforms has found that a range of privacy and security concerns influence decisions to share. For example, one study found that who was requesting information and the purpose for the request had the greatest influence over decisions to disclose or withhold, with the user's actual location and current activity being less important [5]. Relatedly, location information has been identified as the most sensitive element of information shared within social networks [23].

5. TECHNICAL CHALLENGES FOR LOCATION PRIVACY

Below, we briefly discuss the challenges of providing technical support for the expression of location privacy preferences. Location privacy needs to be understandable to technically unsophisticated users ("no surprises"), needs to be implementable within common Web and application development environments, and avoid fictions such as that users read and understand privacy statements. In general, it needs to be easy for applications developers to "do the right thing", given that location-based applications will be created not just by large corporations like Facebook or Google, but also by many individual developers or small teams without a dedicated legal staff, e.g., as part of plug-ins for social network applications or in open-source projects and small web sites.

Discussing location privacy is made more difficult because of the wide range of privacy concerns related to location data and the wide range of location resolution possibilities. First, even systems that do not explicitly deal with geographic coordinates or street addresses reveal location data. For example, every web connection that does not use an anonymizing proxy reveals the user's IP address to the server, which can usually be readily mapped, using widely available commercial tools, to city-level location, and, in some cases, to a much finer granularity. For example, university campuses and larger companies typically have their own IP address ranges.

Thus, to discuss location privacy more systematically, we need to distinguish what kind of information is gathered along with location data, who receives the data and whether the data can be aggregated and combined with other data. The simple transmission of a location may, depending on these circumstances, raise very limited privacy concerns or allow near-perfect personal tracking. As an example, if a user provides his current location to a mapping service, without using a personal login, cookies or similar identifying information, and the mapping service deletes the location immediately after showing the map, the privacy implications are likely limited. However, a very similar service, e.g., for vehicular navigation, could continuously collect and store a user's location data, along with timestamps, and make the location track available to a host of third-party entities, or expose it through a social network service to a large group of individuals. In both cases, the mobile device transmits exactly the same information via the same protocols, yet the threats to the user's privacy are likely to be perceived as much more grave for the navigation service.

Even without identifying information, location tracking can reveal a user's home address and employer, simply by looking for the typical night and day-time locations. This information may be discernible even from information aggregated across time or even if noise is added to the location information, as the noise may average out for locations where users spend a significant amount of time.

From the example, we can note that we need to consider at least three facets of location information: the spatial resolution of the location data, whether the location information can be readily tied to a person (or vehicle), and whether the data that is made available to third parties includes accurate time information. As for all privacy-sensitive data, it matters greatly how long the data is stored and who is given access to the data. For storage, a service could store data, e.g., as part of a service log, on off-line tapes, making it accessible only with manual effort, e.g., after a subpoena.

Users cannot be expected to analyze the fine points of a company's business relationships and system architecture. However, there may be some basic categories that can distinguish services by the degree of potential privacy concerns. One model is to provide basic information to users in a standardized graphical format, similar to how the "Schumer box" summarizes some of the core features of a credit card policy. The two basic categories are the purpose the location information is used for, and how long the information is retained. To identify the purpose, the basic distinction is simple: is the location information used only to provide the service the user requested, such as mapping or presence, or is it also used for other purposes, such as location-specific advertising or user profiling? An important related consideration is whether the information is passed on to other applications affiliated with the service, such as Facebook applications.

The IETF GEOPRIV privacy framework defines a related notion of retransmission to "other parties" and allows location objects to allow or prohibit such retransmission. However, this raises the question of what constitutes another party. In the example of location-based advertising, if a service provider owned an advertising clearinghouse or had only a 49contracted out the delivery of advertisements to others, it may fall into the opposite category. Also, most services routinely use other parties, e.g., web hosting companies or cloud services, as part of providing their primary service. Thus, to avoid leaving implementors at the mercy of angels-on-a-pin lawyering, a clearer definition or alignment of retransmission with the purpose of data collection and use would be beneficial.

For the location data retention time, rough categories such as "below 1 day" may be sufficient, rather than specifying a precise duration. In the opinion of one of the authors, the GEOPRIV model of allowing users to specify the retention time appears less practical, as this makes implementation significantly more complicated. If the data contains the retention duration, the service provider either has to be able to handle per-request retention, which is likely to be impractical, or deal with the case where the requested retention time is below the duration implemented by the service. The service would then have to make sure that it rejects the request, with some kind of sensible error message, but also delete the request from its logs. Such a requirement is likely to be burdensome, inviting either implementors that ignore these corner cases or legal challenges and bad publicity even with innocent mistakes.

Thus, designing suitable privacy mechanisms for location data is more than a protocol or legal exercise, but needs to take into account the realities of web service development and the limited ability and willingness of users to understand the intricacies of web architectures and business relationships.

Third-party "black box" privacy rating of services does not appear to be feasible, so the user has to trust the service provider or possibly a third-party certification agency that can audit the operational practices of the location-based service provider.

As for all personal data, it would seem useful for the target (the entity being tracked) to be able to observe what information has been stored. The benefits of such access need to be weighed against the risks of disclosure to other parties that may obtain the user's credentials, since password recovery and easy-to-guess passwords often make it all to easy to break into web applications. A mechanism to delete all stored data may be useful to limit the damage if the user realizes that he or she has disclosed a particularly sensitive location.

6. CONCLUSIONS

In this paper, we argue that location as a concept adds a new quality to the Web, because it adds an element of immediate and direct physical identifiability that has not been part of the location-unaware Web. While location is the first such concept to "enter the Web", others will follow, such as camera access, streaming audio from microphones, and a plethora of other data that already is available through the *mobile Web* and will grow substantially with the *Web of Things*. The current approach to deal with these changes seems to be rather ad-hoc and also does not take into account some of the efforts and experiences that other communities have already invested in exploring these new areas. We argue that for the Web to better deal with its role as the one information systems that seems to slowly devour all the others, it would be helpful to learn from previous lessons in these areas, instead of starting from scratch and developing ad-hoc solutions. The *W3C Device APIs and Policy Working Group* may provide one existing opportunity to apply this approach, but advancing the Web strategically will require looking beyond the API level, and will require incorporating location together with other new mobile Web concepts. A more strategic way of developing technical and policy solutions in this space could help the Web to develop faster and in a more coherent and predictable way.

7. REFERENCES

[1] Kim Alexander and Keith Mills. Voter Privacy in the Digital Age, May 2004.

[2] Richard Barnes, Matt Lepinski, Alissa Cooper, John B. Morris, Hannes Tschofenig, and Henning Schulzrinne. An Architecture for Location and Location Privacy in Internet Applications. Internet Draft draft-ietf-geopriv-arch-01, October 2009.

[3] David D. Clark, John Wroclawski, Karen R. Sollins, and Robert Braden. Tussle in Cyberspace: Defining Tomorrow's Internet. In *ACM SIGCOMM 2002 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 347–356, Pittsburgh, Pennsylvania, August 2002. ACM Press.

[4] Julie E. Cohen. Structural Rights in Privacy. *University of Chicago Law Review*, 75(1), 2008.

[5] Sunny Consolvo, Ian E. Smith, Tara Matthews, Anthony LaMarca, Jason Tabert, and Pauline Powledge. Location Disclosure to Social Relations: Why, When, & What People Want to Share. In Katz et al. [14], pages 81–90.

[6] Lorrie Faith Cranor. *Web Privacy with P3P*. O'Reilly & Associates, Sebastopol, California, September 2002.

[7] Lorrie Faith Cranor and Joseph M. Reagle. Designing a Social Protocol: Lessons Learned from the Platform for Privacy Preferences. World Wide Web Consortium, Note NOTE-TPRC-970930, November 1997.

[8] Jorge R. Cuellar, John B. Morris, Deirdre K. Mulligan, Jon Peterson, and James M. Polk. Geopriv Requirements. Internet RFC 3693, February 2004.

[9] Nick Doty, Deirdre K. Mulligan, and Erik Wilde. Privacy Issues of the W3C Geolocation API. Technical Report 2010-038, School of Information, UC Berkeley, Berkeley, California, February 2010.

[10] Michelle Engelhardt Danley, Deirdre K. Mulligan, John B. Morris, and Jon Peterson. Threat Analysis of the Geopriv Protocol. Internet RFC 3694, February 2004.

[11] Mary Flanagan, Daniel C. Howe, and Helen Nissenbaum. Embodying Values in Technology: Theory and Practice. In Jeroen van den Hoven and John Weckert, editors, *Information Technology and Moral Philosophy*, chapter 16, pages 322–353. Cambridge University Press, March 2008.

[12] Dominique Guinard and Vlad Trifa. Towards the Web of Things: Web Mashups for Embedded Devices. In *Second Workshop on Mashups, Enterprise Mashups and Lightweight Composition on the Web*, Madrid, Spain, April 2009.

[13] Dominique Guinard, Vlad Trifa, and Erik Wilde. Architecting a Mashable Open World Wide Web of Things. Technical Report 663, Institute for Pervasive Computing, ETH Zürich, Zürich, Switzerland, February 2010.

[14] Irvin R. Katz, Robert L. Mack, Linn Marks, Mary Beth Rosson, and Jakob Nielsen, editors. *CHI '95: ACM Conference on Human Factors and Computing Systems*, Denver, Colorado, May 1995. ACM Press.

[15] Jennifer King and Chris Jay Hoofnagle. A Supermajority of Californians Supports Limits on Law Enforcement Access to Cell Phone Location Information. Technical report, School of Law, UC Berkeley, Berkeley, California, April 2008.

[16] Martin Kofahl and Erik Wilde. Location Concepts for the Web. In Irwin King and Ricardo Baeza-Yates, editors, *Weaving Services and People on the World Wide Web*, pages 147–168. Springer-Verlag, Heidelberg, Germany, August 2009.

[17] David M. Kristol. HTTP Cookies: Standards, Privacy, and Politics. *ACM Transactions on Internet Technology*, 1(2):151–198, November 2001.

[18] Tara Matthews, Jeffrey Pierce, and John Tang. No Smart Phone Is an Island: The Impact of Places, Situations, and Other Devices on Smart Phone Use. Technical Report IBM Research Report RJ10452, IBM, September 2009.

[19] Alexander Mayrhofer and Christian Spanring. A Uniform Resource Identifier for Geographic Locations ('geo' URI). Internet Draft draft-ietf-geopriv-geo-uri-04, November 2009.

[20] John Morris and Alan Davidson. Policy Impact Assessments: Considering the Public Interest in Internet Standards Development, August 2003.

[21] Deirdre K. Mulligan and Aaron Burstein. Implementing Copyright Limitations in Rights Expression Languages. In *2003 ACM Workshop on Digital Rights Management*, volume 2696 of *Lecture Notes in Computer Science*, pages 137–154, Washington, D.C., November 2003. Springer-Verlag.

[22] Helen Nissenbaum. Privacy as Contextual Integrity. *Washington Law Review*, 79(1), 2004.

[23] Sameer Patil and Jennifer Lai. Who Gets to Know What When: Configuring Privacy Permissions in an Awareness Application. In Katz et al. [14], pages 101–110.

[24] Andrei Popescu. Geolocation API Specification. World Wide Web Consortium, Working Draft WD-geolocation-API-20090707, July 2009.

[25] Henning Schulzrinne. The tel URI for Telephone Numbers. Internet RFC 3966, December 2004.

[26] Harry Surden. Structural Rights in Privacy. *SMU Law Review*, 60:1605–1629, 2007.

[27] United States Code. Drivers Privacy Protection Act (DPPA). 18 USC 2721, January 2009.

[28] Erik Wilde and Antti Vähä-Sipilä. URI Scheme for Global System for Mobile Communications (GSM) Short Message Service (SMS). Internet RFC 5724, January 2010.