HTTPa: Accountable HTTP

Oshani Seneviratne and Lalana Kagal
MIT Computer Science and Artificial Intelligence Lab
32 Vassar Street Cambridge MA 02139
{oshani,lkagal}@csail.mit.edu

November 5, 2010

Introduction

Most discussions of Internet privacy, both policy and technology, tend to assume Alan Westin's perspective [20], which defines privacy as the ability for people to determine for themselves "when, how, and to what extent, information about them is communicated to others". However, this focus on controlling information access has been found to be flawed [7]. This year's technology press is filled with announcements by social networking sites about their new privacy controls, i.e. new ways for users to define access rules [18, 22]; followed by embarrassment when the choices prove to be inadequate or too complex for people to deal with [16, 21, 15, 5, 12]. Even when access control systems are successful in restricting access to particular users, they are ineffective as privacy protection for systems like the World Wide Web, where it is easy to copy or aggregate information. These days, it is also possible to infer sensitive information such as social security numbers (SSN) [11], political affiliations [10], and even sexual orientation [6] from publicly available information. Another problem with using up-front access control systems is that it is the users' responsibility to define and maintain their privacy policies in every domain they participate in. Lastly, in a pure access restriction system, those who obtain access to the data, legitimately or not, can use the data without restriction.

Instead of enforcing privacy policies through restricted access, we suggest using "information accountability". Weitzner et al define information accountability in terms of usage–when information has been used, it should be possible to determine whether the usage was appropriate, identify the violator and hold him accountable [19]. Lampson argues that to be practical, accountability needs an ecosystem that makes it easy for senders to become accountable and the receivers to demand it [9]. In our accountability research, we focus on helping users conform to policies by making them aware of the usage restrictions associated with the data [14, 8] and helping them understand the implications of their actions and of violating the policy, and encouraging transparency and accountability in how user data is collected and used.

In this position paper, we discuss our ideas on adding accountability to the *HTTP* protocol level. By adding policy-awareness, negotiation of access and usage restrictions, and logging of the access and intent directly into this protocol, we hope to make it easier for Web users to track how their data was used and identify inappropriate usage.

Web Protocol for Accountability

Having an accountable Web protocol will help alleviate some of the privacy problems we face today with respect to accessing, transferring and reusing Web content. We propose HTTPa as an extension to HTTP to provide end-to-end accountability on the Web. This protocol will allow servers to be held accountable for what they serve, and users to be held accountable for the data transactions they perform on the Web. Further, it is our intention to develop a system similar to Primelife's "D. Dashboard" [13], where users will be able to see who used their data, when, how, and where it was used.

Functionality of HTTPa

There are several key components in the protocol. First, users will need to identify themselves before initiating a Web transaction. Second, for each transaction there will be an "accountability-aware" log record. These log records will include who accessed the data, what their intention of use was, where the data was relayed to, and other such accountability preserving data. Third, data will be served after some negotiation regarding usage restrictions between the server and a user-agent.

Authentication

Since accountability is the main goal of HTTPa, users of the protocol will need to identify themselves for authentication with the data and service providers they access on the Web. The WebID protocol [1] will be used for this. Of course, if someone wishes to use a particular data or service anonymously, especially if the user does not trust the provider, using the standard HTTP will give those users the incognito mode they prefer. However, it is possible that servers will provide less granular information or even no information, if users are unwilling to commit to the HTTPa protocol and consequently are not willing to be held accountable for information misuse.

Provenance

We propose logging on both ends of the transaction: the user as well as the server. Logging the information pertaining to the data transfer is one way of preserving the data provenance and negotiated usage restriction in HTTPa. Currently read-only logs on the Web servers are used mainly for debugging problems on the server or to generate statistics about how websites are accessed. For each HTTP request, the HTTP method, HTTP version of the client and the server, URL of the requested resource, HTTP status code of the response, size of the request and the response messages, timestamp of when the transaction occurred, referrer and user agent header values are logged. In HTTPa we will need additional data fields related to the transaction such as what data was accessed, what was the specified intent, and what were the agreed upon usage restrictions.

We envision logs in HTTPa to (i) be immutable except by protocol components, (ii) be encrypted, (iii) be readable only by trusted parties, and (iv) have all the records pertaining to a particular data usage.

Having a detailed log on the user side will allow the development of usage-aware tools that take advantage of the log to encourage the user to use the data appropriately. When the user tries to reuse data she got during an earlier transaction, the tools will read the log to figure out if the data was retrieved from another server and retrieve the usage restrictions. The tools would then (i) remind the user of the usage restriction associated with the data, (ii) inform the user if he is violating the usage restriction, or (iii) allow the user to only use the usage restriction associated with that data as the intention for the current transaction.

These logs will also be useful in identifying potential misuse of information. When misuse is suspected, in theory, it is possible to find a path from the server providing the misused information to the violator through the set of servers and users who used and shared that information. Instead of expecting complete provenance trails, it might be possible to ask servers/users at each node of the path to prove that they used the data appropriately and to provide a set of servers/users that they shared the information with.

Negotiation of Usage Restrictions

P3P (Platform for Privacy Preferences) protocol [3] was developed at the W3C with the intention of communicating the privacy policies of Web sites to the user-agents who connect with them. A user-agent retrieves a machine readable privacy policy from the Web server and responds appropriately (for e.g. display symbols or prompt the user for action). According to the protocol, it is also possible to build tools that can compare each policy against the user's usage restrictions and assist the user in deciding when to exchange data with the Web sites. However, P3P has several limitations: complicated language to express policies, inability to express preferences on third party data collection, and the inability to specify multiple privacy policies for one Web page are to name a few [2]. These limitations has prevented P3P from mass adoption.

Learning from the limitations of P3P, we have considered two alternative ways of handling negotiation of usage restrictions: (i) usage restrictions can be sent via an HTTP header, and the user agent has to accept that header before reading, transferring or doing any kind of transformation on the data, (ii) data will be encrypted, and the only way to decrypt would be to accept the terms on usage restrictions (similar to the public key infrastructure used in implementing SSL on HTTPS).

For our initial development, we will consider a simple ontology of usage restrictions such as the Respect My Privacy ontology [8]. This will simplify the negotiation between users and servers with respect to usage restrictions and will be similar to the negotiation suggested for location information [4]. The next phase will include more complex usage restrictions that are composed of contextual and domain specific constraints and will possibly require a multi-step negotiation protocol such as [17].

Summary

This protocol will address the limitations of current privacy work and provide the infrastructure to build more privacy-aware systems. We believe that government organizations, academic institutions, and businesses will be the early adopters of an accountable Web protocol within their intranets. On the longer run, in a similar vein in

which the growth of e-commerce Web sites led to the massive adoption of HTTPs, we envision that HTTPa will be accepted by the larger Web community, as privacy problems slowly cripple the growth of the Web.

References

[1] Webid protocol. *ESW Wiki, http://esw.w3.org/WebID*.

[2] Pretty poor privacy: An assessment of p3p and internet privacy. *Electronic Privacy Information Center, http://epic.org/reports/prettypoorprivacy.html*, June 2000.

[3] L. F. Cranor. Web privacy with platform for privacy preferences. *Oreilly Books*, Jan 2002.

[4] N. Doty and E. Wilde. Simple Policy Negotiation for Location Disclosure. In *W3C Workshop on Privacy and data usage control*, 2010.

[5] GigaOm. Is facebook beacon a privacy nightmare? http://gigaom.com/2007/11/06/facebook-beacon-privacy-issues/, 2007.

[6] C. Jernigan and B. Mistree. Gaydar: Facebook friendships reveal sexual orientation. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2611/2302, 2009.

[7] L. Kagal and H. Abelson. Access control is an inadequate framework for privacy protection. In *W3C Privacy Workshop*, 2010.

[8] T. Kang and L. Kagal. Enabling privacy-awareness in social networks. In *Intelligent Information Privacy Management Symposium at the AAAI Spring Symposium 2010*, March 2010.

[9] B. Lampson. Usable security: how to get it. *Communications of the ACM*, Jan 2009.

[10] J. Lindamood, R. Heatherly, M. Kantarcioglu, and B. Thuraisingham. Inferring private information using social network data. In *World Wide Web Conference poster paper*, 2009.

[11] PC World. Researchers expose security flaw in social security numbers. http://www.pcworld.com/article/167975/researchers_expose_security_flaw_in_social_security_numbers.html, 2009.

[12] PC World. Google buzz criticized for disclosing gmail contacts. http://www.pcworld.com/businesscenter/article/189081/google_buzz_criticized_for_disclosing_gmail_contacts.html?tk=rel_news, 2010.

[13] Primelife. D. Dashboard. http://www.primelife.eu/results/opensource/76-dashboard.

[14] O. Seneviratne, L. Kagal, and T. Berners-Lee. Policy aware content reuse on the web. In *ISWC2009 - International Semantic Web Conference*, October 2009.

[15] The Local. Headmaster fired after Facebook pic scandal. http://www.thelocal.se/20148/20090618/, 2009.

[16] UPI. Waitress fired for Facebook comment. http://www.upi.com/Odd_News/2010/05/17/Waitress-fired-for-Facebook-comment/UPI-39861274136251/, 2010.

[17] D. D. Walker, E. G. Mercer, and K. E. Seamons. Or best offer: A privacy policy negotiation protocol. *Policies for Distributed Systems and Networks, IEEE International Workshop on*, 0:173–180, 2008.

[18] Wall Street Journal. Facebook grapples with privacy issues. http://online.wsj.com/article/SB10001424052748704912004575252723109845974.html?mod=WSJ_Tech_LEFTTopNews, 2010.

[19] D. Weitzner, H. Abelson, T. Berners-Lee, C. Hanson, J. Hendler, L. Kagal, D. McGuinness, G. Sussman, and K. K. Waterman. Transparent Accountable Inferencing for Privacy Risk Management. In *AAAI Spring Symposium on The Semantic Web meets eGovernment*, March 2006.

[20] A. Westin. Privacy and freedom (Fifth ed.). New York, U.S.A.: Atheneum, 1968.

[21] Wikipedia. Star wars kid. http://en.wikipedia.org/wiki/Star_Wars_Kid, 2002.

[22] Wired. Facebook debuts simplified privacy settings. http://www.wired.com/epicenter/2010/05/facebook-debuts-simplified-privacy-settings/, 2010.