Usage Restriction Management for Accountable Data Transfer on the Web

Oshani Seneviratne and Lalana Kagal
Decentralized Information Group
MIT Computer Science and Artificial Intelligence Lab
{oshani, lkagal}@csail.mit.edu

Abstract

We describe a novel way of usage management using a infrastructure that enables accountability on the Web at the protocol level. The protocol, HTTPA (Accountable Hyper Text Transfer Protocol), requires the data producer and the data consumer to come to an agreement before the data transfer, enabling both parities will be held accountable for the agreement they had entered into. The data consumer will express the intentions of data access and usage, whereas the data producer will express the usage restrictions on the data. This data transfer is facilitated by a trusted third party "Provenance Controller" in an "intentions and usage restrictions handshake". The sender/data producer will evaluate to what extent the usage restrictions match the data consumer's intentions. If they match, the data consumer is granted access to the data; else she is notified of the mismatched components. This protocol cannot prevent the unauthorized reuse of data, but rather it can be used to develop accountability mechanisms that will identify violators allowing them to be held them accountable for data they inappropriately consumed and served.

1 Introduction

Most discussions of Internet privacy, both policy and technology, tend to assume Alan Westin's perspective [32], 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". This assumes that there are major privacy risks from unauthorized access to information. This focus on controlling information access has been found to be flawed [11]. The reality is that, even when the information is within reasonable bounds of security, it can leak outside these privacy boundaries violating the initial restrictions imposed on the data, as many social media outlets on the Web provide an easy medium for information dissemination at an unprecedented level. The 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 [30, 36]; followed by embarrassment when the choices prove to be inadequate or too complex for people to deal with [28, 34, 25, 9, 22]. For example, Facebook's changes to its privacy settings in spring 2010 made news that highlighted how convoluted their privacy policy has become [8]. Tools such as a "Terms of Service Tracker" [20] have led to visualizations of how Facebook is sharing more private data than ever before [18]. Also, Facebook's Open Graph Protocol's "like" button has led to possible privacy violations ranging from exposure of browsing habits of people on medical sites to pornographic sites being shared with an unanticipated audience [27].

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) [21], political affiliations [16], and even sexual orientation [10] 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.

A pure notice and choice model is also not an adequate framework for privacy protection. The choice to whether opt-in or opt-out becomes meaningless and "user choice" is becoming a way for the industry to shift blame to users whenever a privacy breach happens. Many websites publish privacy policies which are often very verbose, and rarely do users have the time to read them or understand what they really mean. A typical user will click through the privacy policy statements without completely understanding the risks involved. In a pure access restriction system, those who obtain access to the data, legitimately or not, can use the data without restriction. An example for this, is the controversial whistle-blowing site *wikileaks*. This website exposes sensitive data with the aim of making governments and large businesses more transparent and accountable. Their claim is that no-one has been intentionally harmed so far because

of the data published on the site. However, due to the sensitive nature of the data published on the site, it is possible for nations, if not individuals, to get harmed at some level and diplomatic relations to deteriorate. In a recent memo, several U.S. agencies have issued a warning [33] saying that the documents published on the site "does not alter the classified status or automatically result in declassification of the documents". Further, the memo states that "classified information, whether or not already posted on public websites or disclosed to the media remains classified and unauthorized federal employees should not look at leaked classified data". This *usage restriction* is inherently faulty because there can be no enforcement (unless the employees are only accessing the website from their work computers where their web browsing is monitored), nor can employees be held accountable for accepting the restrictions imposed on the sensitive data.

Therefore, instead of enforcing privacy policies through restricted access, which does not seem to work well in the current Web landscape, 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 violators and hold them accountable [31]. In our accountability research, we focus on helping users conform to policies by making them aware of the usage restrictions associated with the data [24, 13] and helping them understand the implications of their actions and of violating the policy, thus encouraging transparency and accountability in how user data is collected and used. 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 [15]. It is our belief that HTTPA will provide this eco-system.

2 Motivating Scenarios

Users are increasingly finding their information such as personal profiles, friends, and interests spread across multiple social networking sites and accessed by all sorts of people, many of whom they did not originally intend to share their data with. As social media is becoming central to many things ranging from recruiting to personal relationships, the ability to grant and restrict access to personal data is becoming critical. The ubiquity of the Web, the ability to connect data from external sites to the social networking sites, and the amount of time people spend interacting with social media are both advancing our freedoms and enabling novel invasions of privacy. It is our belief that users should be aware of and ideally be in control of information about them on the Social Web. In the scenarios described below, we take a policy-centric view on Social Web privacy, where policies capture the permissions such as access control, obligations such as terms-of-use and licensing, and other data-handling settings that allow a user to control their interactions with other users. In particular, policies apply privacy settings to the profile and social media frameworks to consistently manage the user expectations of privacy and other obligations. This allows individuals and businesses on the Social Web to share information without any fear of violating user privacy or any regulations within the purview of the intention of use of their audience. We draw few examples from the Social Web to illustrate the importance of having the protocol described in this paper for transferring private data on the Web. These examples show how the intentions of data access will be matched up with the usage restrictions imposed on the social data of an individual.

In the following scenarios assume that Alice is a user of an imaginary social networking site called 'SocialBook'. Alice communicates with SocialBook using our protocol, and both parties have specified their intensions and usage restrictions using the RMP (Respect My Privacy) ontology [13]. The Provenance Tracker 'TrustMe' is a third party entity trusted by both Alice and SocialBook.

2.1 Upstream Usage Restriction Management

Suppose Alice wants to upload some pictures on SocialBook. The default settings on her smart Web client is set with the usage restriction that any HTTP payload carrying data with MIME type such as 'image', or subtypes such as 'image/[bmp,gif,jpeg,png,x-ico,x-tiff]' will only be posted/uploaded if the recipient acknowledges the full ownership of the content to her. However, it seems that SocialBook has extremely draconian terms of service that if uploaded to SocialBook, the data becomes the property of SocialBook. Alice's client examines these two policies, and informs Alice about the mismatch, which then prompts Alice to either stop posting her pictures or to notify SocialBook for the potential terms of use mismatch. In the latter case, TrustMe gets a notification of the handshake that happened between the parties. If SocialBook decides to modify the terms of use, it will send another request which Alice accepts and the data will be transferred.

2.2 Downstream Usage Restriction Management

Alice has a photo on SocialBook with a usage restriction specifying that the photo cannot be used for any *commercial* purposes. An employee from a large advertising company, Bob, accessed that photo. Bob's smart client confirmed with SocialBook and was logged on TrustMe that the intention of accessing the photo was *non-commercial*, and that he will honor the corresponding usage restriction that

Alice has imposed on the photo. However, few weeks later, Alice found out that Bob had used her photo in an online advertisement in his company. Through her Web client Alice complains to TrustMe by giving the URI of her photo that Bob had allegedly used. Alice in her compliant also says that Bob's advertisement had used her photo, and that it is of *commercial-use*. TrustMe verifies that Bob had accessed the photo by looking up the accountability logs. Then it looks up the original usage restriction that Bob agreed to, verifies that it had indeed violated Alice's terms of use, and sends a takedown request to Bob with a proof detailing the violation.

3 Usage Restriction Management in HTTPA

The following sections illustrate the key protocol components in HTTPA that facilitates usage restriction management.

3.1 Authentication

Authentication is a crucial component in the protocol, not just for access control, but also to find the identity of the users who accessed resources should their owners claim that someone violated their usage restrictions on those resources. Therefore, it is very important in this protocol for the data requestors and data providers to identify themselves before the data transfer.

Since this is a decentralized system, we require a global identity of the entities involved in a transaction. The WebID protocol [1] provides a robust mechanism for authentication in such a setting. An entity that wishes to access a resource using HTTP over TLS has to go through a Verification Agent[1] that performs authentication on the provided WebID credentials and determines if the requestor can have access to a particular resource.

3.2 Usage Restrictions and Intentions Specification Language

Websites publish privacy policies that communicate planned data handling practices, such as rights of the data, intended purposes of collection, and third parties who may have access to the data collected from the users. Users also have complimentary usage restrictions for what their data can and cannot be used for. For our initial implementation we used the RMP ontology[2]. This ontology allows specifying usage restrictions and intentions for:

- No Ownership Transfer (the ownership of the data item is with the data producer and it is non-transferable)
- No Commercial (the owner of this data does not want the information on this profile used for commercial purposes)
- No Depiction (the owner does not want her picture used for any reason and does not want her private information used to identify her in an image)
- No Employment (the owner of this data does not want the information on this profile used for employment purposes)
- No Financial (the owner of this data does not want the information on this profile used for financial decisions)
- No Medical (the owner of this data does not want the information on this profile used for decisions related to medicine or medical care)
- No Insurance (the owner of this data does not want the information on this profile used for decisions related for Insurance purposes)

3.3 Usage Restrictions Management

We have explored two approaches in handling usage restrictions management with the intentions for data access:

1. Usage restrictions and the intentions are sent via HTTP headers. A smart agent at the sending end will assess the compatibility of the usage restrictions with the intention before sending the data to the requestor.
2. Data will be sent encrypted to the recipient without performing any usage restriction and intention matching. Decrypting the data will signal accepting of the usage restrictions, and the Provenance Controller will be notified of the agreement.

We are still at the initial stages of the implementation of the HTTPA, and we have a prototype implementation for usage restriction management using method 1 outlined above.

3.4 Handshake

In this protocol, we break away from the traditional client-sever model of HTTP transactions, and allow clients to act as servers, and vice versa. The Sender (server/data provider) conveys usage restrictions, and the Receiver (client/data consumer) notifies her intentions on the data. Figure 1 shows a sequence diagram showing how HTTPA

[1]The Verification Agent is typically a Web server, but may also be a peer on a peer-to-peer network.

[2]The RMP (Respect My Privacy) ontology is available at `http://dig.csail.mit.edu/2008/02/rmp/rmp-schema.n3`

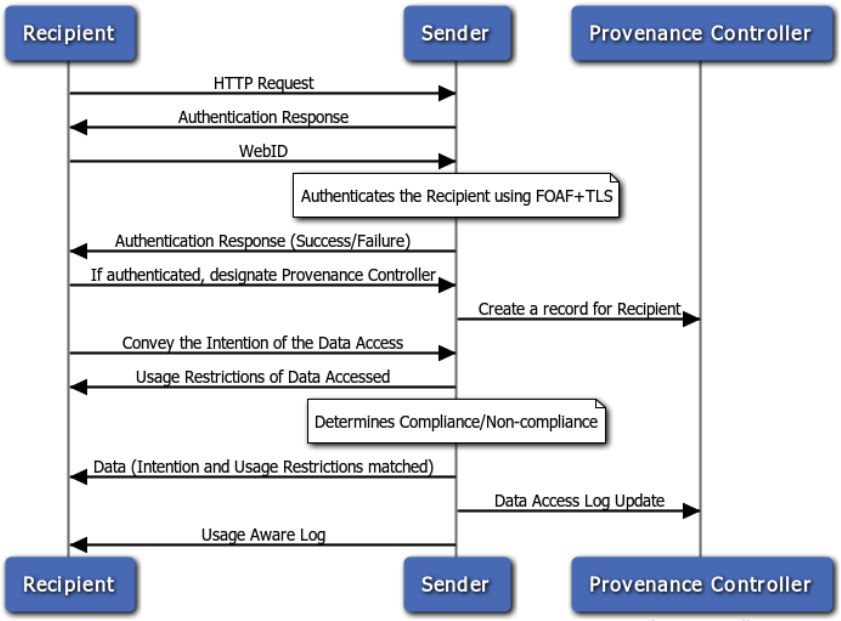

Figure 1. Sequence of transactions in HTTPA. Usage Restriction Management is carried out when the Recipient (data consumer) conveys the intention for data access and the Sender (data producer) matches those intentions with the usage restrictions of the data that was accessed.

facilitates a handshake between the sender and the recipient of the data before the actual transfer.

In the current implementation, we define two HTTP Headers: 'X-UsageRestrictions' and 'X-Intentions' for the usage restrictions and the intentions respectively. More than one usage restrictions/intentions can be sent in these headers by separating them with commas in the value field.

For example, in the scenario listed in Section 2.1, Alice's Web client sends the following HTTP request to SocialBook:

```
POST /path-of-updatable-document HTTP/1.1
Content-Type: image
X-UsageRestrictions: http://dig.
csail.mit.edu/2008/02/rmp/rmp-schema#
No-Ownership-Transfer
```

SocialBook responds to this with the following HTTP response:

```
HTTP/1.1 412 Precondition Failed
X-Intentions: http://dig.csail.
mit.edu/2008/02/rmp/rmp-schema#
Ownership-Transfer
```

Alice's Web client reads the X-Intentions HTTP header, and notifies Alice about the mismatch. If Alice wishes to renegotiate, she has two options:

1. She can instruct her Web client to send the following HTTP request renouncing her usage restriction:

   ```
   POST /path-of-updatable-document HTTP/1.1
   Content-Type: image
   X-UsageRestrictions: http://
   dig.csail.mit.edu/2008/02/rmp/
   rmp-schema#Ownership-Transfer
   ```

   Since this usage restriction matches perfectly with SocialBook's intention for the data, the server returns an HTTP 200 response, and accepts her picture to be published on the website.

2. She can ask the SocialBook to honor her usage restriction and send the data by sending the following HTTP request with the X-Negotiate header.

   ```
   POST /path-of-updatable-document HTTP/1.1
   Content-Type: image
   X-Negotiate: http://dig.csail.
   mit.edu/2008/02/rmp/rmp-schema#
   No-Ownership-Transfer
   ```

   SocialBook upon receiving the X-Negotiate header will acknowledge the request by either: agreeing with

Alice's usage restriction and sending the data with an *HTTP 200 OK* response, or ignore the request by sending an *HTTP 204 No Content* response.

We can imagine a similar handshake for Bob's accessing Alice's picture for a *non-commercial* use in the scenario outlined in Section 2.2. Alice and Bob will be both using the usage restriction/intention `http://dig.csail.mit.edu/2008/02/rmp/rmp-schema#No-Commercial` which leads to the successful data transfer.

3.5 Provenance Controllers

Provenance Controllers are essentially special Web servers that are delegated to handle logging to enable provenance in HTTPA transactions. They are somewhat analogous in concept to Certificate Authorities used in Public Key Infrastructure because they are trusted by both parties involved in the data transfer. They also search through and reason over the logs to determine the identity of entities that accessed the data, in case the data providers claim that someone has violated their usage restrictions.

In cases where there is a usage restrictions violation, the data owner can complain to the provenance controller she had designated for the data transfer by giving the URI of the original content (URI-O), and the URI of the content in question (URI-Q). The provenance controller then sends a 'provenance-tracking' request to the server that hosts URI-Q. This server will then provide a list of provenance controllers it had dealt with to the requesting provenance controller. It will then propagate requests to each of the provenance controllers in the list that it obtained, to determine the identity of the person who put up that content at URI-Q.

3.6 Logging

Currently read-only logs[3] on Web servers are used for debugging problems on the server or to generate statistics about how websites are accessed. In our prototype implementation, the following logs are maintained by various protocol components in addition to these server-side debug logs:

- **Accountability Logs**: These are created by the Provenance Controller for every HTTPA transaction between a Sender and a Receiver. Accountability Logs have several characteristics: they are immutable except by protocol components, encrypted, secure, readable only by trusted parties involved in the HTTPA transaction, and have all the records pertaining to a particular data transfer and usage such as what data was accessed, the specified intent of access, and the agreed upon usage restrictions.
- **Usage-Aware Logs**: These are sent to the receiver's client by the Provenance Controller. Smart clients at the receiver's end are able to understand the usage restrictions associated with the data, and warn the user about a potential misuse. For example, the client can warn the user if she is copying and pasting parts from a resource that was originally marked as non transferable.
- **Data Provenance Logs**: These are created by smart clients on the data receiver's end. The smart clients help facilitate the user in creating a remix from several different resources gathered from the Web, and during this process, it constructs a provenance trail with the URIs of the HTTP resources used in the remix. When the user posts the remixed content on a server, the smart client can optionally send the header *X-Meta* with a link to the log file associated with the content that is being posted. The server upon receiving this information will pass on to the provenance controller associated with this specific HTTP transaction.

3.7 Accountability Checking

If a user finds that she was wronged because someone else misused her data by violating the usage restrictions associated with the data, she can take recourse by producing a provenance trail with the help of the provenance controller. For instance, in the scenario given in Section 2.2, Alice can complain to her trusted provenance controller through her client that URI-Q has material from URI-O, and that it violated her usage restriction on *non-commercial* use. Alice's client will produce a 'complaint' in RDF using the Notation 3 syntax [26] as follows:

```
<URI-O> rmp:usage_restriction rmp:non-commercial-use.
<URI-Q> rmp:contains <URI-O>,
 rmp:inappropriately-used-for rmp:commercial-use.
```

This complaint will be communicated to the provenance controller (PC-A) that Alice dealt with, when uploading the content on URI-O. PC-A will poll the server hosting URI-Q to give the list of provenance controllers it had dealt with. Upon receiving the list of Provenance Controllers, PC-A will poll each of the provenance controllers in the list for the identity of the entity that posted content at URI-Q. Additionally, PC-A can request the composition of the content at URI-Q provided that Bob had sent that using the optional *X-Meta* header. Then, based on the evidence available, PC-A will execute the following rule written in the AIR policy

[3]In Apache2, 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 for each HTTP request.

language [12] to determine if there has been any usage restrictions violation, and if so, by whom.

```
@forAll :VIOLATOR, :ORIGIN, :DEST, :CONTENT, :USAGE,
 :CURR_USE.

:UsageRestrictionCheck a air:Policy;
   air:rule :CheckInAppropriateUsage.

:CheckInAppropriateUsage a air:Belief-rule;
     air:if { :CONTENT rmp:usage_restriction :USAGE.
                :DEST rmp:use :CURR_USE;
                          dc:author :VIOLATOR;
                          rmp:contains :ORIGIN.
                :USAGE owl:differentFrom :CURR_USE.
     };
    air:then [ air:assert
             {:VIOLATOR air:not-compliant-with :USAGE}].
```

Figure 2. AIR policy to determine whether the destination's current use is different from the intended use of the content, and if so, the author of that destination content is in violation.

Once the violator has been identified as Bob, PC-A can send a notification to him detailing Alice's complaint. We are also working on language components to describe what Alice can request Bob to do if PC-A determined that Bob had indeed violated Alice's usage restrictions. For example, Alice can request Bob to takedown the content in URI-Q, or Alice can request compensation.

4 Related Work

Various machine readable approaches to describing privacy policies have been proposed over many years. P3P (Platform for Privacy Preferences) protocol [4] was developed at the W3C with the intention of communicating the privacy policies of websites to the user-agents who connect with them. The recommendation allows website operators to express their data collection, use, sharing, and retention practices in a machine-readable format. A user-agent can retrieve a machine readable privacy policy from the Web server and respond appropriately (for e.g. display symbols or prompt the user for action). However, P3P has several limitations: a complicated language to express policies, inability to express preferences on third party data collection, and to specify multiple privacy policies for one Web page [2]. These limitations have prevented P3P from wide adoption. Unlike in P3P, both parties have a say in the data transfer in our protocol. Also, our work attempts to bring down the complexity barrier by making the usage restrictions and the intentions expression simpler with the help of smart clients.

The W3C POWDER (Protocol for Web Description Resources) language provides a mechanism for describing groups of resources by essentially grouping URIs and linking these groups of URIs to a group of common XML statements regarding topics like authentication [3]. While more generic than P3P, it was aimed at similar privacy use-cases such as privacy descriptions for child protection. While it is interesting that it describes groups of URIs rather than single URIs, it is seen as complex and has failed to gain deployment for the same reasons as P3P. Unlike in POWDER, our proposal provides de-referenceable URIs that points to the usage restrictions and the intentions. Also, since the usage restrictions are expressed in RDF, it is more expressive than the POWDER descriptions in XML.

FTC endorsed a 'Do not Track proposal' [7] recently to facilitate consumer choice about online tracking, and there are already several implementations that support this proposal. One of the most compelling technical implementations describes sending the user's intention of not to track online browsing behavior in an HTTP header [17]. Although this approach works for this specific use case, it seems very limited for general purpose usage restrictions matching with intentions. Also, the communication described in their proposal through HTTP Headers is monodirectional, whereas our protocol allows bi-directional communication enabling both parities to engage in a dialogue.

Specific to geo location data, several proposals on how to negotiate privacy policies have emerged within the IETF and the W3C recently. IETF's GeoPriv proposal [5] attempts to put privacy policies in the hands of users instead of services, where a user transmits her own privacy preferences about how her data should be used, while the websites are bound by their market or legal obligations to respect those preferences. W3C's Geolocation API [23] also advocates websites to disclose their data usage practices to the user, although it is rarely practiced by most websites that implement the API [6]. The Simple Policy Negotiation for Location Disclosure proposal [35] describes a system that lets a user have a dialogue with a website that uses her location data before disclosure. Their proposal has many similarities to our accountable data transfer protocol, such as implementing a simple standard for transmitting policy information just-in-time. However, their domain is limited to geolocation data, and the their standard does not handle provenance tracking.

Mozilla Privacy Icons takes a simple icon-based approach inspired by the Creative Commons [19]. Instead of specifying every possible type of privacy and data-handling scenario, they specify only a few common privacy scenarios that users can encounter such as information sharing, storage, monetization, deletion and contact/notification. As online businesses are looking for ways to build trust and manage consumer expectations through transparency, choice, and accountability, these privacy icons can help online businesses achieve that. The icons are designed to be easy to

use and be understood by ordinary end-users. But because there is no incentive for sites that violate user privacy to label themselves as such, it would be up to the browser, or a browser app, to automatically label such sites. Also, users do not ordinarily notice an icon by its absence but only by its presence. Therefore the browser/app should detect the absence of the privacy icons to notify users they have entered a site where their privacy and usage restrictions could be violated. Although it does not address every possible scenario, this approach manages to defeat the complexity barrier of rule-based approaches. In Section 5 we illustrate how we plan to integrate the the Mozilla Privacy Icons to our protocol.

5 Future Work

In our current implementation, the user has to specify usage restrictions and intentions, either per site, per groups of sites, or just have a default setting applied for all data transfers. This model is not foolproof, as it is possible for someone to access one resource from the a site for one purpose, and access another resource from the same site for another purpose. It is also not feasible to require the user to specify the usage restrictions and the intentions for each and every HTTP transaction she performs. We are planning on exploring ways in which it would be possible to capture a user's intentions by her web browsing behaviors as described in [14].

Although we have implemented all the protocol components, we are yet to finish the implementation of the user interface components, including the 'smart client'. We are also considering integrating with the Mozilla Privacy Icons project once they have evaluated how online privacy related text can be modularized with the list of usage restrictions/intentions settings that need to be visualized, and determine the 'on' and 'off' states of privacy settings based on the user's contextual information. This will be integrated with the proposed smart client that keeps track of the provenance of the constituents of a particular resource, and communicates the data provenance log to the server that is accepting the data.

The next phase of our protocol will include more complex usage restrictions that are composed of contextual and domain specific constraints. We are exploring multi-step negotiation protocols such as the one described in the "Or Best Offer" privacy policy negotiation protocol [29] to handle more complex scenarios. On the other extreme, we can achieve a robust access control mechanism by simplifying the protocol, and this is also something we are aiming to work on in the future.

6 Conclusion

The protocol described in this paper addresses the limitations of current privacy work and provides the infrastructure to build more privacy-aware systems. The requestor, on data access will convey what her intention for the data access is. The data provider will determine the compliance/non-compliance of the intention sent by the requestor with the usage restrictions associated with the resources that are being accessed. Their negotiation is being logged by a trusted third party called 'Provenance Trackers' to ensure accountability. If usage restrictions are compliant with the intentions, the data access request will be successful. If it is non-compliant, an explanation as to why the data cannot be transferred will be conveyed to the requestor. The recipient of the data will be held accountable for the usage restrictions she accepted upon the data transfer. In other words, recipients cannot argue after the fact that they did not know the expectations of the data server: for retention or for use of information. Similarly, users cannot claim after the fact that the data server was deceptive or that they had not been informed. This enables market and regulatory forces to punish users who misuse data. We believe that government organizations, academic institutions, and businesses will be the early adopters of this accountable Web protocol with usage restriction management 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. *WebID 1.0 - Web Identification and Discovery,* `http://getwebid.org/spec/drafts/ED-webid-20100809/index.html`.

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

[3] P. Archer, K. Smith, and A. Perego. Protocol for Web Description Resources (POWDER): Description Resources. `http://www.w3.org/TR/powder-dr`.

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

[5] J. R. Cuellar, J. B. Morris, D. K. Mulligan, J. Peterson, and J. M. Polk. Geopriv Requirements. Internet RFC 3693. `http://www.ietf.org/rfc/rfc3693.txt`.

[6] N. Doty and E. Wilde. Geolocation privacy and application platforms. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS*, SPRINGL '10, pages 65–69, New York, NY, USA, 2010. ACM.

[7] Federal Trade Commission Staff Report. Protecting consumer privacy in an era of rapid change - a proposed framework for businesses and policymakers. http://www.ftc.gov/os/2010/12/101201privacyreport.pdf, 2010.

[8] G. Gates. Facebook privacy: A bewildering tangle of options. http://www.nytimes.com/interactive/2010/05/12/business/facebook-privacy.html.

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

[10] 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.

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

[12] L. Kagal, I. Jacobi, and A. Khandelwal. Gasping for air: Why we need linked rules and justifications on the semantic web. In *Under review at the World Wide Web Conference 2011*, 2010.

[13] 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.

[14] R. Kumar and A. Tomkins. A characterization of online browsing behavior. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 561–570, New York, NY, USA, 2010. ACM.

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

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

[17] J. Mayer and A. Narayanan. Do Not Track - Universal Web Tracking Opt Out. http://donottrack.us.

[18] M. McKeon. Facebook privacy vizualizer. http://mattmckeon.com/facebook-privacy/.

[19] Mozilla. Privacy Icons. https://wiki.mozilla.org/Drumbeat/Challenges/Privacy_Icons.

[20] T. T. of Service Tracker (A project of the Electronic Frontier Foundation). Facebook privacy policy. http://www.tosback.org/policy.php?pid=39.

[21] 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.

[22] 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.

[23] A. Popescu. Geolocation API Specification. http://www.w3.org/TR/geolocation-API.

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

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

[26] Tim Berners-Lee and Dan Connolly and Lalana Kagal and Jim Hendler and Yosi Schraf. N3Logic: A Logical Framework for the World Wide Web. *Journal of Theory and Practice of Logic Programming (TPLP), Special Issue on Logic Programming and the Web*, 2008.

[27] M. Tuffield. Nhs.uk allowing google, facebook, and others to track you. http://mmt.me.uk/blog/2010/11/21/nhs-and-tracking/.

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

[29] 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.

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

[31] 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.

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

[33] Wikileaks Access Warning. http://edition.cnn.com/2010/US/12/03/wikileaks.access.warning.

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

[35] E. Wilde. Simple policy negotiation for location disclosure. *w3.org*.

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

A Appendix: Namespaces Used

```
rmp: http://dig.csail.mit.edu/2008/02/rmp/rmp-schema#.
air: http://dig.csail.mit.edu/TAMI/2007/amord/air#.
dc: http://purl.org/dc/terms/.
```