Augmenting the Web with Accountability

Oshani Seneviratne
PhD Candidate, Decentralized Information Group, MIT

Advisor: Tim Berners-Lee

PhD Symposium at WWW2012
Agenda

• Motivation
• Proposed Approach
• Methodology & Evaluation
• Challenges
Agenda

• Motivation
• Proposed Approach
• Methodology & Evaluation
• Challenges
Accountability on the Web?
Accountability on the Web?

- Controlling access has been the de-facto method for protecting information in computer systems.
Accountability on the Web?

• Controlling access has been the de-facto method for protecting information in computer systems

• But information can leak even though there are strict access control mechanisms
Accountability on the Web?

- Controlling access has been the de-facto method for protecting information in computer systems.
- But information can leak even though there are strict access control mechanisms.
- We need accountability to **supplement** access control on the Web [Weitzner et al, 2008]
Unintended Consequences from Private Information Leakages on the Web
Unintended Consequences from Private Information Leakages on the Web
Copyright Infringements
Copyright Infringements
Goals
Goals

- Enable provenance of web resources
Goals

• Enable **provenance** of web resources
• Prevent **unintended consequences** from misusing information on the web
Goals

• Enable **provenance** of web resources

• Prevent **unintended consequences** from misusing information on the web

• Enable reuse of web content in a **progressive policy aware** manner
Agenda

• Motivation
• Proposed Approach
• Methodology & Evaluation
• Challenges
HTTPA
(HTTP with Accountability)
HTTPA
(HTTP with Accountability)

- Secure Protocol
HTTPA
(HTTP with Accountability)

- Secure Protocol
- Has Auditing Capability
HTTPPA
(HTTPP with Accountability)

- Secure Protocol
- Has Auditing Capability
- Extensible Privacy and Reuse Restrictions
HTTPA in a Nutshell
HTTPPA in a Nutshell

Client - Information Transfer - Server
HTTPA in a Nutshell

Client --> Information Transfer --> Server

Provenance Tracker
Network
HTTPA in a Nutshell

Client

Information Transfer

Server

Provenance Tracker

Network

Provenance Trails
HTTPA in a Nutshell

Client

Information Transfer

Server

Logs

Logs

Provenance Tracker

Network

Provenance Trails

Friday, April 20, 2012
HTTPA Method Sequence
HTTPPA Method Sequence

Consumer

Provider

Consumer

Provider
HTTPA Method Sequence

Consumer

HTTPA <METHOD> URI

Provider

Consumer

Provider
HTTPA Method Sequence

Consumer

HTTPA <METHOD> URI

Authentication Request

Provider

Consumer

Provider
HTTPA Method Sequence

HTTPA <METHOD> URI

Authentication Request

WebID Protocol

Consumer

Provider

Consumer

Provider
HTTPA Method Sequence

Consumer -> Provider

HTTPA <METHOD> URI

Authentication Request

WebID Protocol

Usage Restrictions

Consumer -> Provider

Friday, April 20, 2012
HTTPA Method Sequence

Consumer -> Provider

HTTPA <METHOD> URI

Authentication Request

WebID Protocol

Usage Restrictions

Agreement & Intentions

Consumer -> Provider
HTTPA Method Sequence

Consumer → Provider

HTTPA <METHOD> URI

Authentication Request

WebID Protocol

Usage Restrictions

Agreement & Intentions

Consumer

Provider

Provenance Tracker Network

Provenance Tracker Network
HTTPA Method Sequence

Consumer

Provider

HTTPA <METHOD> URI

Authentication Request

WebID Protocol

Usage Restrictions

Agreement & Intentions

Web Resource

Accountability Log

Acknowledgement

Provenance Tracker Network

Consumer

Provider

Provenance Tracker Network

Friday, April 20, 2012
Protocol Components

- Authentication
- Usage Restriction Specification
- Provenance Tracker Network
- Smart Clients
- Accountability Checking
Authentication
Authentication

• For:
  - Authenticity
  - Access Control
  - Auditing
Authentication

- For:
  - Authenticity
  - Access Control
  - Auditing
- Implemented using the WebIDs
Usage Restrictions
Usage Restrictions

• Default Terms Supported:
  - Read
  - Read+Embed
  - Read+Copy
Usage Restrictions

- Default Terms Supported:
  - Read
  - Read+Embed
  - Read+Copy

- The current implementation handles the Respect My Privacy (RMP) ontology
Provenance Tracker Network (PTN)
Provenance Tracker Network (PTN)

- Maintains the accountability logs for web resources
Provenance Tracker Network (PTN)

- Maintains the accountability logs for web resources
- Trusted Service
Provenance Tracker Network (PTN)

- Maintains the accountability logs for web resources
- Trusted Service
- Experimental PTN deployed on PlanetLab
Provenance Tracker Network (PTN)

- Maintains the accountability logs for web resources
- Trusted Service
- Experimental PTN deployed on PlanetLab
Provenance Tracker Network (PTN)

- Maintains the accountability logs for web resources
- Trusted Service
- Experimental PTN deployed on PlanetLab
Smart Clients
Smart Clients

- Manages the user’s intentions for accessing generic web resources
Smart Clients

• Manages the user’s intentions for accessing generic web resources

• Displays the Usage Restrictions set on the content access
Smart Clients

- Manages the user’s intentions for accessing generic web resources
- Displays the Usage Restrictions set on the content access
- Preserves the provenance and rights information
Smart Clients

- Manages the user’s intentions for accessing generic web resources
- Displays the Usage Restrictions set on the content access
- Preserves the provenance and rights information
- Implemented as browser based plugins
Smart Clients

- Manages the user’s intentions for accessing generic web resources
- Displays the Usage Restrictions set on the content access
- Preserves the provenance and rights information
- Implemented as browser based plugins
Smart Clients

• Manages the user’s intentions for accessing generic web resources

• Displays the Usage Restrictions set on the content access

• Preserves the provenance and rights information

• Implemented as browser based plugins
Accountability
Checking
Accountability Checking

• A rights owner for some web content can request the PTN for the following information:
  - Usage of their web resources
  - Adherence to their usage restriction terms
Accountability Checking

- A rights owner for some web content can request the PTN for the following information:
  - Usage of their web resources
  - Adherence to their usage restriction terms

- PTNs provide an ‘audit trail’ with:
  - URIs of original resources
  - The identity of the violator
Agenda

• Motivation
• Proposed Approach
• Methodology & Evaluation
• Challenges
Studies on Content Reuse on the Web
Studies on Content Reuse on the Web

• Flickr study revealed CC license violations ranging from 78% - 94% on Flickr Images [Seneviratne, Kagal and Berners-Lee ISWC 2009]
Studies on Content Reuse on the Web

- Flickr study revealed CC license violations ranging from 78% - 94% on Flickr Images [Seneviratne, Kagal and Berners-Lee ISWC 2009]

Sample 1 (67 Pages, 426 Images)
Sample 2 (70 Pages, 341 Images)
Sample 3 (70 Pages, 466 Images)
Studies on Content Reuse on the Web

- Flickr study revealed CC license violations ranging from 78% - 94% on Flickr Images [Seneviratne, Kagal and Berners-Lee ISWC 2009]

- Survey study on user generated content websites [Seneviratne, Monroy-Hernandez WebSci 2010]
Implementation
Implementation

- Usage Restriction Management on Web Resources
  [Seneviratne, Kagal IEEE Policy 2011]
Implementation

- Usage Restriction Management on Web Resources
  [Seneviratne, Kagal IEEE Policy 2011]

- Initial PTN deployment on PlanetLab
Implementation

• Usage Restriction Management on Web Resources [Seneviratne, Kagal IEEE Policy 2011]

• Initial PTN deployment on PlanetLab

• Ongoing protocol implementation work (see http://dig.csail.mit.edu/hg/httpa)
Planned HTTPA Trials
Planned HTTPA Trials

- Library of Congress information resource management of their “Born Digital” collection
Planned HTTPA Trials

- Library of Congress information resource management of their “Born Digital” collection
- HTTPA to reward long tail content creators with the **TipSy** and **Emancipay** projects
Planned HTTPA Trials

- Library of Congress information resource management of their “Born Digital” collection
- HTTPA to reward long tail content creators with the TipSy and Emancipay projects
- Prototype Social Networking Site
Agenda

• Motivation
• Proposed Approach
• Methodology & Evaluation
• Challenges
Privacy
Privacy

• Use of HTTPA does not mean giving up pseudonymity or anonymity
Privacy

- Use of HTTPA does not mean giving up pseudonymity or anonymity
  - **Anonymizing Agents** will preserve the identity of the user
Privacy

• Use of HTTPA does not mean giving up pseudonymity or anonymity
  - Anonymizing Agents will preserve the identity of the user

• Tracking provenance: Is this the end of privacy?
Privacy

• Use of HTTPA does not mean giving up pseudonymity or anonymity
  - Anonymizing Agents will preserve the identity of the user

• Tracking provenance: Is this the end of privacy?
  - Accountability will ensure the preservation of privacy in the long run
Scaling Up
Scaling Up

- Hierarchical PTN Infrastructure
Scaling Up

• Hierarchical PTN Infrastructure

• Organizations can run their own PTN, much like mail servers, certificate authorities, etc
Thank You!

Please send your comments to:
oshani@mit.edu
Appendix
WebID Protocol

1. TLS setup
2. cert request
3. x509 certificate
4. HTTP response in TLS
5. HTTPS GET
6. identity query
7. authorization query
8. exponent
9. modulus

Friday, April 20, 2012
Related Work

- P3P [Cranor-2002]
- Distributed Usage Control [Kumari-2010]
- Project DReaM [Sun-2009]