Identifying and Preventing Conditions for Web Privacy Leakage

Craig E. Wills
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA USA

W3C Workshop on Web Tracking and User Privacy

April, 2011

It's Not Just Tracking

...but also leakage of private information via a range of first-party sites (employment, news, travel, entertainment...) utilizing registered user accounts (not just OSNs!) to third parties.

Leakage may be explicit (intentional?) or implicit (inadvertent?). Leakage may be included as part of privacy policy or it may not.

Bottom line, it's not just tracking of user behavior by third-parties, but the capability to link this behavior and bits of private information about users.

How Does Leakage Occur?

- Via information embedded in URLs and leaked via `Referer` header.

  GET http://tracker.thirdparty.com/params...
  Referer: http://www.firstparty.com/...zip=12201...

- Via page titles

  GET http://tracker.thirdparty.com/...title=John Doe profile...
  Referer: http://www.firstparty.com/profile/123456789...

- Via information stored in first-party domain cookies, which are passed to *hidden* third parties.

  GET http://thirdparty.firstparty.com/...
  Referer: http://www.firstparty.com/...
  Cookie: ...e=jdoe@email.com&f=John&l=Doe...

- Via population of third-party requests

  GET http://tracker.thirdparty.com/...age=30&gender=M&zip=12201...
  Referer: http://www.firstparty.com/...

How Can Leakage be Prevented?

Limited means for users beyond blocking third-party requests.

An opt-out cookie does not prevent leakage. A "Do Not Track" header does not prevent leakage.

Best done by first parties to negate a condition causing leakage.

- not embedding private info in URLs—using POST to send data rather than a paramaterized GET request.
- Not putting info in page titles where JavaScript APIs can access the info.
- Not embedding private info in domain-wide first-party cookies so that hidden-third parties on these domains get access to the info.
- Not allowing third parties access to the private information provided by users to first parties.