Minutes of the IAB/W3C/ISOC/MIT Internet Privacy Workshop

December 8-9, 2010
Boston, MIT

Attendees

Fu-Ming Shih, MIT
Ian Jacobi, MIT
Steve Woodrow, MIT
Nick Mathewson, The Tor Project
Peter Eckersley, Electronic Frontier Foundation
John Klensin, IAB
Oliver Hanka, Technical University Munich
Alan Mislove, Northeastern University
Ashkan Soltani, FTC
Sam Hartman, Painless Security
Kevin Trilli, TRUSTe
Dorothy Gellert, InterDigital
Aaron Falk, Raytheon - BBN Technologies
Sean Turner, IECA
Wei-Yeh Lee, NAVTEQ
Chad McClung, The Boeing Company
Jan Seedorf, NEC
Dave Crocker, Brandenburg InternetWorking
Lorrie Cranor, Carnegie Mellon University
Noah Mendelsohn, W3C TAG Chair
Stefan Winter, RESTENA
Craig Wittenberg, Microsoft
Bernard Aboba, IAB/Microsoft
Heather West, Google
Blaine Cook, British Telecom
Kasey Chappelle, Vodafone Group
Russ Housley, IETF Chair/Vigil Security, LLC
Daniel Appelquist, Vodafone R&D
Olaf Kolkman, IAB Chair
Jon Peterson, IAB/NeuStar, Inc.
Balachander Krishnamurthy, AT&T Labs--Research
Marc Linsner, Cisco Systems
Jorge Cuellar, Siemens AG
Arvind Narayanan, Stanford University
Eric Rescorla, Skype
Cullen Jennings, Cisco
Christine Runnegar, Internet Society
Alissa Cooper, Center for Democracy & Technology
Jim Fenton, Cisco
Oshani Seneviratne, MIT
Lalana Kagal, MIT
Fred Carter, Information & Privacy Commissioner of Ontario, Canada
Frederick Hirsch, Nokia
Benjamin Heitmann, DERI, NUI Galway, Ireland
John Linn, RSA, The Security Division of EMC
Paul Trevithick, Azigo
Ari Schwartz, National Institute of Standards and Technology

David Evans, University of Cambridge
Nick Doty, UC Berkeley, School of Information
Sharon Paradesi, MIT
Jonathan Mayer, Stanford University
David Maher, Intertrust
Brett McDowell, PayPal
Leucio Antonio Cutillo, Eurecom
Susan Landau, Radcliffe Institute for Advanced Study, Harvard University
Dave Crocker, Brandenburg InternetWorking
Christopher Soghoian, FTC In-house Technologist, Center for Applied Cybersecurity Research, Indiana University
Trent Adams, Internet Society
Thomas Roessler, W3C
Karen O'Donoghue, ISOC
Hannes Tschofenig, IAB/Nokia Siemens Networks
Lucy Elizabeth Lynch, Internet Society
Karen Sollins, MIT
Tim Berners-Lee, W3C

Tuesday, December 7
===================
5:00pm - Welcome Reception
Location: MIT Stata Center 4th Floor, R&D Pub
(Sponsored by ISOC, and Nokia Siemens Networks)

Wednesday, December 8
=====================
8:00am - 9:00am - Breakfast Coffee
9:00am - 9:30am - Opening Remarks and Logistics, Goal Setting

Speakers:
- Opening Remarks and Logistics: Karen Sollins (including brief intro of the workshop organizers and the persons in the room; name and affiliation)
- Goal Setting: Hannes Tschofenig

Karen Sollins (MIT): Opening Remarks

Hannes Tschofenig (IAB): Output and Goals

Meeting rules: No Chatham House Rules -- participants are free to blog, tweet, etc. during the event.

There will be no breakout sessions. We will move up the stack as the day progresses: we will start with discussions about Network-level Privacy, then move to the Browser-level. Then, we will discuss how to get privacy considered by standards developers/implementers given what we learned about getting security into the picture.

Workshop Outputs

- Position Papers: http://www.iab.org/about/workshops/privacy/papers/Privacy_Workshop_Papers.zip
- Presentations: http://www.iab.org/about/workshops/privacy/slides/slides-all.zip
- Press Release: http://www.iab.org/about/workshops/privacy/press_release.pdf
- Meeting Minutes: Ed. this writeup
- Meeting Report(s): to-be-done

- Potential standards actions by IETF & W3C

Process for Producing Outputs

- IPR: IETF Note Well does not apply.
- Position Papers
  - Let us know if you do not want to get your position paper published.
  - Deadline: End of the workshop.
- Press release
  - We were approached by the press about the workshop; they are not invited.
  - We will compile a brief high level statement by the end of the workshop.
- Meeting Minutes
  - We will take meeting minutes. Karen found a few scribes among her students.
  - We will distribute the meeting minutes after the workshop to all workshop participants for review.
  - Let us know if you do not want to get quoted or attributed.
  - Tell us now or review the meeting minutes.
- Meeting Report (s)
  - Higher level summary, to be developed based on the minutes.
- Potential standards actions by IETF & W3C
  - Developed independently by both organizations, on their own timelines and their own rules.

Workshop Goals:

How do we design systems so that they respect privacy?

- What does "privacy by design" mean for technical people?
- What is the role of policy makers in standardization?
- What guidance can we give to specification authors and implementers?
- How do protocol and architectural design decisions help to shape the privacy properties of the overall system?

Is there specific work the W3C / IETF can/should do?

Reference Development Lifecycle:

- Research
- Standardization
- System design / implementation
- Deployment

Session Structure:

- How do systems work today?
- How should they work (desired state)?
- Is there a way to get to the desired state?

Questions we will run into:

- When you say "privacy", what do you mean?
- What is the threat model?
- How do we get the incentives right?
- How much do we rely on support from the legal framework?

- What can we reasonably expect of users?
- What is our technical approach for dealing with privacy beyond data minimization?

Comments
Alissa Cooper, CDT:

- In addition to the development lifecycle and the actors involved, there are also user advocates, policy makers, etc. [My recollection may be off, but I thought my point here was that we needed to consider a variety of deployment scenarios, including shared-computing situations such as family and library computers, that introduce different privacy challenges from single-user scenarios. -Alissa]
- While the IAB/IETF/W3C can't police protocol implementations, they need to consider what incentives can be provided to coordinate the chain of actors.

Frederick Hirsch/Nokia: Another two goals:

- How does privacy design play out when we have a number of standards organizations with different timelines, different objectives, etc.?
- Also, how do we ensure system-level/system-wide privacy by design when a system is commonly composed of a number of components?

David Crocker: It is nice to see the scope kept fairly narrow. That said, the focus on browsing doesn't necessarily encompass distributed processing, web applications, etc.
Karen Sollins/MIT: What about "architecture", between the "research" and "standards" development blocks? Research often contains architectural work, but also there are architectural considerations throughout.

9:45am - 10:15am - Privacy at IETF/W3C

How did we deal with privacy in the absence of a commonly understood approach for dealing with privacy? Short examples from the IETF and the W3C.

Jon Peterson (IAB): IETF Privacy Presentation

- Protocols assume architectures
  - Ideally, these protocols should be useful in a variety of architectures.
  - However, certain protocols are not useful in some.
- IETF doesn't mandate implementation style and deployment characteristics, but does constrain them in various ways.
  - Example: DNS was designed to have a single root.
- Ostensibly, the network intermediaries makes simple forwarding decisions, doesn’t inspect or log packets in any deeper semantics
  - Today, we have plenty of reason to fear otherwise
- What does an application need to share to get a service delivered, and with whom?
  - Intermediation
    - SIP as defined by RFC 3261 uses proxies to route requests based purely on the Request-URI.
    - However, "intermediaries" (such as Back-to-Back User Agents/Session Border Controllers) inspect many other elements of requests (and this "application-layer deep packet inspection" is legitimately valuable). It might be valuable to secure SIP to prevent oversharing (e.g. S/MIME support in RFC 3261), but at the same time the optimization benefits disincentivize this (e.g. RFC 3261 S/MIME was rarely implemented).
  - How can SIP share with intermediaries only the information they need to do their job? (RFC

3323 is a start)

- How do we get other protocol designs to learn from this experience?
    - ALTO (ongoing right now): tradeoff between user privacy interests and network service provider privacy interests
    - How can the user share enough with the network for it to be useful and vice-versa?

"Hemispheres" in ALTO

H1: P2P users want to do ALTO, but do not want to disclose information about the overlay
H2: Network operators want to do ALTO, but do not want to disclose information about the network topology and state.

How do we bring them together?

IPv6 Privacy Addresses

- In IPv6 stateless address autoconfiguration the interface identifier was constructed based on the interface MAC address
- This raised privacy concerns
- RFC 4941 supported a dynamically generated IPv6 interface identifier

Peter Eckersley (Jabber Room): speaking of IPv6, does anyone have a sense of the extent to which RFC 4941 is actually being used in today's IPv6 deployments?
Jim Fenton (Jabber Room): I asked someone about that the other day and Windows Vista/Windows 7 has RFC 4941 support.
[Editor: Summary of RFC 4941 support is available here: http://kb.wisc.edu/ns/page.php?id=13736]
Dave Crocker: concerning adoption of RFC 4941, I'll make a small plug for recording an adoption assertion in <http://trac.tools.ietf.org/misc/outcomes/>. This makes the assertion public and revisable.

Questions

- Threat model: who are we attempting to hide the address from?
    - 1. ISP eavesdroppers (where?): if an attacker can snoop the Neighbor Discovery (ND) exchange or has access to the ND tables in the router (e.g. an admin with SNMP access), it can obtain the binding of the IPv6 privacy address to the MAC address, a stable long-term identifier.
    - 2. Other communication partners (eg. websites)? Those attackers have access to techniques for identifying users other than IP address, including cookies, fingerprinting, etc.
    - 3. If the government has access to the ISP records, see #1. If it has access to data from communications partners, see #2.
- Example from the emergency services field mentioned. To obtain emergency services over VOIP, it is helpful for a device to convey its location to the Public Service Answering Point (PSAP) via SIP. Without location, police/fire/medical will have a difficult time responding. The same mechanisms that allow ISPs to track users are used to provide location for emergency services and to deal with certain security attacks (botnets).
    - Location is often obtained from the access network, which needs to map the IP address to a location (e.g. via a wiremap).
    - To map an IPv6 privacy address to a location, the access provider may have to do more work than if the IPv6 address were assigned via DHCPv6, but the mapping of the IPv6 address to a location is still possible.
- Classic “presence” problem
    - I might want to share different presence information with my friend than with my boss (RFC 2778).

        - Per-user/per-"friend" granularity of presence data allows you to control how much presence information you share and with whom.
        - Had we defined “presence” as a unique rather than a potentially manifold property, however, would this be possible?
- Some presence architectures admit of only one view of presence, which is either shared with a particular recipient or not.
    - We layer our basic architecture for geolocation privacy on top of this (RFC 4119).
- However, just because you choose to share information selectively, what about those you shared it with?
    - Policy framework in GEOPRIV enables expressing usage preferences about retention, redistribution, and so on
    - Geolocation information might be valuable for emergency services, but could also be (ab)used for location-based advertising. How do we manage this tradeoff?

How do we learn from these experiences and generalize to other protocol designs?

What we need (and want from the workshop)

- Guidance to authors of protocol specifications on at least four fronts:
    - How do we build privacy threat models?
    - How do we design protocols that do not fall into obvious privacy traps?
    - What are some common ways around traps that you can't get out of?
    - How do we document traps that we don't know how to get out of?

Reference: http://tools.ietf.org/html/draft-morris-privacy-considerations

Discussion

Jon Peterson: One of the major problems addressed in GEOPRIV is "why/how is this information being released"? rather than just a binary state where privacy is "on" or "off". Explicit specification of allowable usage purposes can be helpful.

Alissa Cooper/ CDT: RE: ALTO, corporations don't really have "privacy" concerns, but do have business interests in protecting topology information. Should this workshop consider the privacy of individuals, and not be overly concerned about the privacy of corporations? Can we scope down privacy to mean "personal privacy"?

?: Yes, but we need to be realistic about what the actors think about and what they're willing to do in order to gain mutual benefit.

Dave Crocker: Perhaps the term “confidentiality” would be more effectively employed with corporations. I would also like to explore issues around “disclosure” in addition to “privacy”

Dave Crocker (Jabber Room): I'm wondering whether it will help us to gravitate towards an action label, rather than a state label, for the focus of the workshop? "Privacy" is a state; worse it seems to have very different scope to different participants. One term that was suggested was "confidentiality" and another was "disclosure".

John Linn: Protocol-visible privacy aspects are just the tip of the iceberg -- they provide guidance to system implementers, but don't ensure system behavior or full implementation behavior. Can we encourage data minimization throughout the entire process?

Karen Sollins / MIT: How do we minimize information exposure when exchanging information between parties? How do we allow exposure to be negotiated?

Christopher Soghoian: Most people don't have problem with 911 knowing your location when you call 911. The concern is covert information collection by governments when emergency service is not requested.

Bernard Aboba: By example, what could have been said early in IPv6 design that would have improved the privacy impact? What can we ask protocol designers to consider?

Jim Fenton/Cisco: Re: IPv6 and anticipating future use. We can't expect people to think of every possible

privacy issue that will emerge in future, but having this discussion opens up the review process and may indeed catch more issues than we currently are catching.
Noah Mendelsohn/Chair of W3C TAG: Sometimes we utilize mechanisms for purposes for which they weren't designed -- i.e. Google Analytics puts Javascript on a page that makes an HTTP GET, which provides more information to Google about users that one might not want to disclose -- HTTP was never intended for tracking users.
Dave Crocker: We need an engineering methodology for dealing with privacy...

Thomas Roessler (W3C): W3C Privacy Presentation

1990s: Platform for Privacy Preferences Project (P3P)

- W3C P3P 1.0 A New Standard in Online Privacy (http://www.w3.org/P3P/)
- Lawrence Lessig: "P3P products will allow users to be informed of site practices (in both machine and user readable formats), to delegate decisions to their computer when appropriate, and to allow users to tailor their relationships to specific sites"
- Approach: site shows policy, browser vets it, user can override
  - Policy includes: purpose, recipient, retension
- Deployment: IE6: cookie blocking tied to P3P policy
  - Since pages that placed cookies needed a policy, sites using cookies needed to include **some** policy.
    - Many "cut and paste" policies (even though example pages said "of course, don't just use the coe above as-is")
  - Leon, Cranor, McDonald, McGuire 2010: Large number of wrong policies, but no enforcement. (see http://www.cylab.cmu.edu/research/techreports/2010/tr_cylab10014.html)

P3P was focused on explaining policies for explicitly-shared information, and used a technical approach with legal backing. It was an attempt to build global consensus.
Lesson: global consensus among poicy-makers necessary, but not sufficient for success.

Policy expression in a distributed system has a complex incentive structure -- we may not like contextual advertising, but it may very well pay for the content we browse.
Different policies must make a difference in the user experience. Policy makers must have a clear understanding of how potential policies impact the user.

Another example: In-browser Geolocation API:

- JavaScript Geolocation API results in a browser-driven confirmation window allowing geolocation data to be sent to the server.
- Focuses on access control and transparency
- But API is silent on many privacy issues: retention, secondary use, or data minimization are left to the Web application
- Results: Doty, Mulligan and Wilde (2010): http://www.ischool.berkeley.edu/research/publications/2010/mulligan/privacy
- Compare this with GEOPRIV, where "sticky policies" travel with location data (implying policy-enabled data processing)

It's expected that these dilemmas will be repeated again in other APIs.
Old joke: On the Internet, noone knows you're a dog. But with devices such as microphones and webcams, is this still true?
User behavior monitored against the user's will: is this inherent to the technology? a usability failure in the client? regulatory failure?
We're talking about privacy considerations for a generative platform.

- Conceptual framework?
- Ecosystem interactions?

Consider: the "evercookie": user behavior monitored against their will (see http://www.schneier.com/blog/archives/2010/09/evercookies.html)

- *evercookie* is a javascript API available that produces extremely persistent cookies in a browser. Its goal is to identify a client even after they've removed standard cookies, Flash cookies (Local Shared Objects or LSOs), and others. evercookie accomplishes this by storing the cookie data in several types of storage mechanisms that are available on the local browser. Additionally, if evercookie has found the user has removed any of the types of cookies in question, it recreates them using each mechanism available.
- We need to be careful to not declare the technology involved inherently bad -- browser HTTP caching is extremely important, but it was used against user intentions in this case.
- Can we fix this, or is it inherent in the technology?
- Was it a usability failure in the client (e.g. Flash cookies being hidden)?
- Was it a regulatory problem?
- Should we constrain the application platform?

Again - we are thinking about a generative platform here -- how much do we/should we constrain it? By comparison, what are the privacy recommendations of C/libc?

Discussion
Kasey Chappelle/Vodafone: P3P failed partly because it was before its time -- the legal and regulatory standard at the time was to require disclosure of "terms and conditions". With no other guidance, companies will pack everything into policies as dictated by their lawyers. Privacy policies are moving from "disclosing everything" to "disclosing things that matter". The FTC report (http://www.ftc.gov/opa/2010/12/privacyreport.shtm) demonstrates that we're looking at a sea change.
Christopher Soghoian: The substantive, useful information is never disclosed in privacy policies -- i.e. data retention lifetimes. So the reality is that we have big complex policies that don't disclose "things that matter". Lots of controversy about transparency of privacy policies (e.g. Verizon and Facebook, see http://www.nytimes.com/2007/10/16/business/16phone.html , http://mashable.com/2010/08/25/facebook-privacy-infographic/)
Lorrie Faith Cranor: P3P designers acknowledged expression of "things that mattered" (things like secondary use) but companies didn't want to disclose.
Frederick Hirsch: The workshop shouldn't get stuck assuming that “privacy” is all about what users want; corporations do have legitimate requirements, too (like minimizing liability).
Ari Schwartz/NIST: Not everyone in industry, not all nations/cultures, etc. have the same incentives and it's important to keep this in mind as we go forward.
???: Companies would see value in clearly articulated cause/effect models
???: Even when something is simple and clear, there is a lot of complexity in the meaning that is hard to convey.
John Klensin: P3P is much too complicated for end-users, and put implementers in a difficult position in designing the user experience to present to users. Overly complex solutions result in poor implementations: implementations that are hard to deploy, provide incorrect information, are easily subverted.
Eric Rescorla: The HIPAA privacy policy from your doctor's office seems pretty clear in that they'll disclose the information in a number of vague situations -- i.e. in response to insurance companies or law enforcement subpoenas. However, if you want to know the reasons that the information will be disclosed to law enforcement, then you have to know the case law! The patient has no idea whether the doctor will even fight to not disclose. We need to distinguish between norms, desires and exceptions.
Thomas Roessler/W3C: We should distinguish between things:

- That will necessarily occur during service delivery (i.e. the postman looking at the package address);

- That are controllable (i.e. sending questionable advertising to your mailbox);
- That are exceptions to normal behavior (i.e. law enforcement intercepting mail)

Frederick Hirsch: In P3P, everything's mediated through the browser, but does it necessarily have to be? Could it be mediated at the application layer, by the service provider? Who can take responsibility? Who is responsible? Who enforces this?
Olaf Kolkman: To keep focused, how would you phrase a successful outcome of this workshop when reporting back to your organization?
Thomas Roessler: A successful outcome would provide answers to the following questions:

- What is a conceptual framework for dealing with privacy issues?
- How are other standards organizations dealing with these problems?
- How can the W3C standards development process more effectively focus on privacy issues?

10:15am - 12:00am - Network-Level Privacy Issues

Covered topics:
IP address hiding, mobility, onion routing/TOR,
location information by operators,
network access AAA

Basic structure:
1) This is how systems works now.
2) How should they work (desired state)?
3) Is there a a way to get to the desired state?
(min. 45 min for item #3)

Example to stimulate discussion is Tor
(Nick Mathewson)

Moderator: Lucy Lynch, Eric Rescorla

Network-Level Privacy Issues

Moderator: Lucy Lynch (see Lucy's introductory slide deck)
Definition of privacy: Sharing (data) in an explicit contrxt with an expectation of scope.

- Privacy is not anonymity (though it is a tool of privacy)

Explicit Context Needs:

- Clearly defined practices
- Identification of extended use (of data)

Some of the privacy-issues are artifacts of how a distributed network functions
Goal: participants should encounter useful topics and take them back to consider how to approach solutions

Nick Matthewson: Tor Project
Goal: Making communications unlinkable and untraceable
Bathing suit example (i.e. comparing choices between changing on the beach vs in a changing room) ← **Ask Nick for this**
Privacy is not just limited sharing, but is the non-sharing of what you do not chose to share
The underlying technology for Tor is Onion Routing – routing traffic through multiple servers, each stripping

out more of the identifying data
The Tor stakeholders may be broader than you'd think:

- Obvious: privacy-respecting people
- Not obvious: Governments operating within countries outside their control (e.g. embassies)

About 200k users so far

- Tor network grows when articles appear about Tor
- Users "opt in" to use of Tor. This makes it easy to detect Tor traffic leaving the network.
- It is "opt out" at the exit end. Most places don't opt-out deliberately, although they might catch it as spam/bot detections.

How do you measure how anonymous a system is?

- Look at it from an attacker's point of view: How many individuals might have been the originator of an action?
- Usability (even if the application is simple) requires numerous other people to be using it, too
    - “Anonymity loves company”: as more people use Tor, potential effectiveness improves

Limitations of Tor

- If an adversary can snoop both at the entrance and exit of the network, then it can correlate the end points.
- Long-term identifiers cause issues. This includes addresses, cookies, fingerprinting.
- Problems with IP-based geolocation (you bring up the Google homepage over Tor and it is in the wrong language).
    - If the user wants to be geolocated, then Tor isn't right for them.
- Performance (round-trip time can be increased significantly). We can't reduce performance by a 50 percent!
    - The users find the system is slow, but the people who stay are the people who can tolerate the privacy/speed tradeoff (TLS over TCP),
    - Tor isn't based on P2P technology. Techniques for doing this (P2P) don't work for anonymity networks, since we can't play tit for tat while being anonymous.
- Epistemic attacks can happen when people know about the network. Tor has to do source routing.<-
- Interested in anonymity/technical solutions? References:
    - Privacy Enhancing Technology Symposium. People wanting to develop technology to detect who can learn what. Mailing list: http://lists.links.org/pipermail/pet/
    - Nicholas Hopper: http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/h/Hopper:Nicholas.html
    - Nikita Borisov: http://hatswitch.org/~nikita/publications.html
    - Fitz…Hoffman? "Pseudoanonymity paper"] **<- Need help on this**
- Need more people – the fewer the users, they become more easily identified rather than more anonymous
- Applications are difficult to configure to really offer truly anonymous browsing. Web browsers don't normally offer what we need.
    - Have implemented browser-side anonymity only for Firefox (with Torbutton plugin)
    - So many browser features need to be turned off that many sites are unusable.
- There are inherent issues relating to network path selection being an identifying characteristic.
- We'd like to have an official statement from the W3C saying: “Fingerprinting should be avoidable.”
    - Can we also please break linkability?

Participant Comments

[I find it a bit odd that sometimes commenters are identified by name, sometimes their affiliations (either SDO or company) are given, and sometimes no information is given at all. It seems like whatever the citation style is, it should be uniform, otherwise it's difficult to follow who was speaking.-Alissa]
Sean Turner: Should the IETF explore and evaluate standardizing anonymizing components that Tor can leverage?
Anonymity is not a competitive advantage (e.g. customers will give up data in exchange for goods)
Are people more willing to participate when they know they're anonymous?

- Answer: Yes, in specific circumstances (e.g. spousal abuse, political commentary)

Is Google large enough to be the omniscient observer who could re-collate a user's traffic?

- Answer: Possibly, depending on the Google's presence in the countries of origin.

What is the current state of level of service?

- Answer: bad. Need more nodes and some changes should be made to the protocol-level

Reference (from Jabber): http://www.thoughtcrime.org/software/sslstrip/

- How will the move toward Routing PKI affect Tor? To prevent spammers stealing address blocks temporarily (and injection of bogus routing advertisements), we are moving toward a "titling" system for address blocks - RPKI.   RPKI eliminates pseudonymity (an address block owner can't advertise someone else's block on the network). Answer: Tor exit nodes need to demonstrate legitimate ownership of address blocks.

What browser features need to be turned off for Tor to provide anonymity?

- Need to override large portions of JavaScript
- Need to block nearly all other browser plugins
- Every tracking component needs to be subverted

Which apps does Tor support?

- If application appears to work great, but you don't know you're not anonymous, then it's a bad application. Is it just as fast as your real network? Yeah… That's a problem (Tor probably isn't running).
- Need to answer what applications we KNOW Tor works with. Requirements:
  - Support SOCKS4/SOCKS5 proxy support, patched library calls can do it, so that connect does a SOCKS5 connect.
  - DNS is a pain.
  - Long-term identifiers are a problem, fingerprinting, are problems.

What could the technical community do to help solve these issues?

- Would like to see people talk to the Torbutton maintainer about issues, since he lists threat models, possible attacks, and what changes are needed.
- Would like a recognition in principle that these are issues. We have a hard time getting people to consider the attacks as "attacks that matter". Fingerprinting, for example. No one thought this would be something to be resisted. If someone said fingerprinting should be avoided, we could get standard bodies to think that people care about this.

What would be the consequence of losing the ability to use fingerprinting as a mechanism for security?

- Until enough people see value in being anonymous, there aren't any economic drivers to come up with alternate approaches
- Fingerprinting is one example of the trade-offs that enable one set of functionality at the sacrifice of anonymity
- Reference (from Jabber): if anyone wants to know about fingerprinting, https://panopticlick.eff.org/browser-uniqueness.pdf
- Comment (from Jabber): that was discussed in that Gartner report on using other persistent tracking mechanisms for fraud detection: http://www.gartner.com/DisplayDocument?ref=clientFriendlyUrl&id=1297620 - for other reasons though, I know many security folks that don't log into Paypal and banks in the same browser that they use for Facebook or to browse the web. This is an attempt to reduce the exposure of XSS and XSRF attacks.

Are there more privacy-friendly solutions for things that rely on fingerprinting? If not, can there be? How can we influence their development?
If the major browser vendors simply refuse to support anonymity, are there any options?
Does it look like we're going down a path where each application will need it's own Tor-like solutions to work?

- Reference: GSMA Mobile Privacy Initiative
  - See http://www.gsmworld.com/newsroom/press-releases/2009/2532.htm
- Are there network-layer solutions that might fall short of perfect anonymity but are still useful in expressing privacy policies (and enforce them)?
- How do you avoid over-specification when trying to point out privacy pitfalls in a specification?

12:00am - 1:00pm - Lunch
1:00pm - 2:45pm - Browser-Level Privacy Issues

Covered topics:
Cookies, private browsing, advanced web API

Basic structure:
1) This is how systems works now.
2) How should they work (desired state)?

Example to stimulate discussion is private browsing
(Christopher Soghoian)

Moderator: Dan Appelquist & Bernard Aboba

Dan Applequist: Some guidelines:

- Let's focus on “Web clients” rather than just “Browsers” so as to support all applications running on the web (not all of which are Browsers).
- We will focus on "Web clients", and exclude the server side.
- Discussion will also focus more on commercial (mis) use cases than on scenarios involving government as an adversary. The goal of the session is to talk about “threat models” within specific contexts.

Christopher Soghoian: “Private Browsing”
Consumers expect features such as “private browsing” to protect them. However, the evidence suggests that consumers to do not understand what threats they are being protected against. Privacy protection natively

supported within browsers is important, since add-ons are not widely used by the general population.

Privacy features native in the browser include:

Private Browsing Mode

- Opera was first, all major browsers include it now.
- This feature protects you against others who access your browser's history (i.e. casual investigators).
- It does not protect against network snooping by an ISP or employer (traffic is not encrypted); malicious code (e.g. spyware); remote tracking or forensic analysis

Evidence indicates that users do not understand what threats "Private Browsing" protects/does-not-protected against:

- Even though Mozilla states their Private Browsing doesn't protect against an employer, usage indicates they're most often used at work during lunch, during which employees are likely using corporate networks, possibly to do things that their employers would not approve of (e.g. surfing porn).
- Reality: even though everyone avoids talking about it, Private Browsing is really “porn mode”.Honest conversations of the usage mode would help.

Claims made by the browser manufacturers:

- "Leave no trace/Like you were never there": a deceptive claim (see http://www.switched.com/2010/07/28/private-browsing-sessions-not-so-private/)
- Google's claims for Chrome are the clearest
- IE/Firefox claims aren't as clear (what's a cookie?)

Third-party Cookie-blocking

- Browsers all differ in how they block third-party cookies
- IE, Chrome, Safari only block new third-party cookies.
- Facebook is an example of a 1st party cookie that is still accessible via private browsing mode

Referrer header blocking

- Sites easily learn where you've been based on the Referrer header; for example, search queries are leaked. This is not communicated to users.
- Information provided by the Referrer header is seen as a feature to the CEO community.
- It is not clear that users view this as a "feature" or a "bug" (privacy violation).
- Referrer header can be disabled (but not easily):
  - Mozilla – can be disabled in a preference setting
  - Chrome – can be disabled via command line

Participant Comments

Is it a reasonable assumption that if the user sees their name in association with a “Like Button”, they know they're being tracked?

- Facebook users probably have no idea what the Like button is doing.
- Users have a limited understanding because they don't know they're being tracked.
- Half of Gmail users didn't even know that they were being targeted based on their e-mails.
- Users don't really understand the trade-offs and need help

    - The ability to get users to understand something is not straight-forward
    - There is a long-term negotiation between users and developers when new security enhancements break things they use.
    - Do we need focus group testing of consumer reactions? We need to be able to make this understandable to the average person.
    - What if people did understand? Would this change their behavior? ("Give me a little information and I'll give you something nice!")
    - People don't understand the role of their ISP (let alone what Tor does).
    - There is today a lack of acknowledgement that there's an economic tradeoff here. The user is getting access to "free stuff" by trading personal information, but they don't understand that this trade is being made, and a (potentially "non-free") alternative is not explicitly offered.

Is the term “Private Browsing” useful, or does it need to be changed?

- It is more effective to talk to users about “goals” (i.e. what the user wants to do) than “mechanisms” (e.g. activate this mode).
- When turning on “Private Mode”, users may expect that it turns on something like Tor.
    - The list of functionality that needs to be disabled/circumvented in Tor mode (and the number of applications that won't work without this functionality) make it clear how difficult it is to attempt to provide untraceability/unlinkability (and how limiting this is).
- How do we explore the context of privacy with respect to:
    - what (doing something naughty)
    - with whom (who do you want to keep information from)

Is privacy like being pregnant (i.e. you either are private or not), or is there a gradient?

- Privacy is multi-variant – there are different degrees... but when considering each context, it IS a binary (yes | no)
- Are there possibilities to explore something more like a three-level gradient of choices for “Private Mode”?
    - Low – Protect against casual interlopers
    - Medium – Protect against cross-server tracking
    - High – Protect against government-level inspection

Considering the “Porn Mode” idea, are there multiple contexts with separate issues to be addressed?

- Context: (example) spousal abuse support
- Classes of issues:
    - Protecting against other people using your computer (private browsing helps)
    - Protecting against services tracking them (private browsing doesn't help)
    - Protecting against government (private browsing doesn't help)

Looking for statistics about usage of “Private Mode”

- 4th most widely used menu option in Mozilla (based on heat-map analysis) ← **need to verify**

Core issue is that users don't really understand just how prevalent tracking is:

- So they focus on the “uncomfortable moment” when unexpected URL-autocompletion occurs indicating their spouse is visiting naughty sites
- It might help to use examples for the options, rather than descriptions of what they do

When working on a disclosure/consent based model, it is necessary to ensure users truly understand (aren't just exposed to) the issues.

- This is more complex, even a small bit of (acceptable) sharing of info can be (silently) aggregated into an (unacceptable) amount of shared information

It is important to consider economic models as part of this conversation (don't forget who's paying many of the bills for all the “free” content):
No existing anonymity solution will survive inspection by the government
Can I use two different browsers to maintain effective separation?
While we may be able to control what happens on servers and clients, but not as much about what happens in the middle.
Scope of privacy issues include confidentiality, cache handling, history, fingerprinting, cookies, etc.

- Blowing out cookies is a nuclear option that many people avoid as it's not fine-tuned enough.

Could we explore models supporting groups of:

- Sites I trust
- Sites I don't trust

… so that decisions could be made based on these distinctions?

What role, if any, might there be for the W3C to answer the questions being brought up?
Exploring the attack models

- Considering the feature sets
- On a technical level, offer a description of what various modes do in different “privacy modes”

… but how does the W3C encourage adoption of these ideas?

A goal out of this session should be learning what pieces of the platform exist that can be further explored

- Explore interaction between security/privacy, feature, and commerce models (cross-reference to Lucy's Venn diagram)

What do we need to provide privacy?

- Anonymized IP addresses.
- no visible fingerprint
- the public mode isn't mapped to private mode
- cookies aren't trackable

… there are a lot more flexible ways within the HTML / JavaScript / Embeds, etc that (when removed) will significantly limit utility.

Consider the picking of fonts as the fingerprint method:

- The vast majority of sites don't make use of the list of fonts
- It is a reasonable idea to explore a small subset (e.g. serif, sans-serif, mono-space) which is less identifiable.

Is there a way to externalize the costs to offset profiling

- e.g. the BBC license fee and a model for paying for “public good” sites
- Labeling the use and intent of data and then enforce it (e.g. cookies labeled “marketing” so the browser can act accordingly)
    - Problem is that self-labeling doesn't work (e.g. one of the P3P problems)
    - Will the FTC (or similar) really enforce this?

· Reference: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1568385

We need to understand the difference between “good” and “bad” actors

- Good = labeling might work
- Bad = labeling won't work (they're going to do what they want)

Collaborative labeling may at least help add clarity and (while not solving the problem) move he ball forward

Designing interfaces around this are nearly impossible, the only solution is within the realm of social sanctioning

- Consider example of learning when it's appropriate to be naked in the front yard (OK for young kids, not OK when they get older)

Suggest that we, as technologists, set aside the human interface issues (knowing they're important), but looking at issues like “minimal disclosure” and similar that we can solve in order to make some progress

Trying to identify the technical building blocks absent understanding of the other aspects (e.g. policy, human factors, etc.) may lead us down a wrong path.

- A major issue in browsing mode is service-level correlation
- We need to understand the difference between “proactive” and “retroactive” actions

2:45pm - 3:00pm - Coffee Break
3:00pm - 4:00pm - Browser-Level Privacy Issues (cont.)

3) Is there a a way to get to the desired state?

Questions:

* What new mechanisms are needed?

* What are the implications for existing and
under development IETF/W3C protocol mechanisms?

“Do Not Track” General Discussion

European POV
Regulators haven't gotten to the same point as US and Canada

- Paranoia about cookies in general resulting in consent required to set them in some contexts
- Implementation complexity. Reference: ePrivacy Directive – http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:EN:NOT
- In current state, companies following the US lead can assume they'll largely be in compliance with European regulations

- Heavy focus on disclosure and consent (i.e. notice and checkboxes)

Canadian POV

- Very broad concepts based on fair information practices principles
- No specific legislation, it's very broad
- Nothing specific about concepts similar to “Do Not Track” in the US

The term “Do Not Track” is over-loaded – the term encompasses a number of proposals and disparate functionality. We need clarity around the various proposals and what they accomplish.

Context of tracking is important in the discussion (sometimes tracking is OK, others it isn't)
Mozilla considerations:

- All 3rd party cookies become session-only cookies
- Double-keying cookies between 1st and 3rd parties

Considering two paths:

- a mode which makes changes (allowing users selecting the mode to trade degradation of experience for privacy)
- changes to entire model

Consider the implications of a “logged in” state in which users explicitly share with sites related to that federation with specific privacy rules
How do you explore alternate economics supported by behavioral marketing without tracking you

- e.g. a “Personal Preference Map”

Compare behavioral tracking with Spam

- Legitimate email marketeers are legislatively powerful
- Minor differences between legitimate and “bad” ones

Remember there are legitimate reasons for tracking you

- 3rd parties are often contracted to perform 1st party services (e.g. analytics)
- Simple customization
- Shopping systems (especially complex ones)

Consider the differences between “personally-identifiable data” and “aggregated data”

- e.g. “I don't care who is reading my blog, I just want to know how many people read it”

Part of this is empowering the internal conversation within a company: “Are we complying with a user's request not to be tracked?”

Is it deceptive (even absent of the implementation of “Do Not Track”) not to clear out cookies after explicitly logging out of a site?

- … not necessarily... Consider storing your IdP option so when you return it's easy to log back in.

“Do Not Track” Approaches

Three proposals discussed, which do not appear to be mutually exclusive:

“Opt Out” of ad tracking:

- Opt-out is with each tracking service
- Opting out places a cookie stating you've opted out
- Clearing cookies clears the opt-out cookies
- Opt-out cookies often expire on their own
  - TACO (Targeted Advertising Cookie Opt-Out) Add-in by Christopher Soghoian maintains opt-out cookies (see http://www.abine.com/more/privacyalerts.php)

“Do Not Track” proposal from Stanford:

- Involves an HTTP header placed within a Request
- It's not blocking lists or registries
- They talk about “tracking” as “data retention”
- Supports behavioral advertising without tracking
  - Adnostic browser plugin builds a local aggregation of your browsing rather than focusing on server-level tracking

“Do Not Track” Proposal by the EFF, CDT, etc.

- Reference: http://www.worldprivacyforum.org/pdf/ConsumerProtections_FTC_ConsensusDoc_Final_s.pdf
- Registry-driven list of domain names
- Model was to create the list, then encourage regulation and browser support
- Not an “add blocking” proposal
- Encouraged ad companies to serve ads from specific servers separate from their analytics servers so the ads aren't blocked, just the tracking

What is the difference between 1st party and 3rd party cookies? Are the distinctions understandable to users and enforceable?

- 1st party personalization is generally acceptable
- Generally speaking, 1st party cookies are set by domains clearly understood as being visited by the user

What about the right to be anonymous? Is there a role for anonymity in "do not track"?
How do we consider tracking in the context of smart devices (e.g. BlueRay players)?

What about differentiating between different classes of use of the data?

- So some can be collected while other uses would prevent tracking based on user choices

There is a huge difference between being tracked for:

- advertising
- providing improved services
- detecting fraud
- building a profile that can be used against you

There are more issues about tracking beyond behavioral ad targeting (e.g. search queries to improve results, bid tracking to prevent fraud)

What will be the economic impact of ad engines not being able to track users and build better advertising?

- Worst case = totally closed subscriptions
- Off-set subscriptions for those willing to pay not to be tracked

Tracking can be useful, but must result in effective improvements over non-tracking alternatives
Are there any non-browser tracking issues?

- For example, what if a web service call included a “do not track” header...?
- Stanford proposal looking to extend into the mobile space

A good model to follow is: Conservative in what you send, Liberal in what you receive

4:00pm - 4:30pm - Lessons learned from Security

Explain how we introduced security
into the standardization process.
Questions to the group:
1) Can we re-apply the procedure for privacy?
2) If yes, how can we do it?

Discussion starter: Russ Housley

Russ Housley / IETF / Defining Privacy
Talking about a definition of privacy in the RFC, a discussion of the security considerations section of all RFCs and how successful this has been, and how this might be used to provide a similar feature for privacy issues in Internet standards.
"Privacy Considerations" sections would be a place in each RFC where the authors are required to make explicit the privacy implications or assumptions of a protocol. Also, unlike security considerations, not every protocol will necessarily be privacy-preserving. I.e. for TCP you could enumerate all of the things (i.e. timers, options, etc.) that could leak information about you, but not necessarily provide any solutions because there may not be solutions.
Sam Hartman: BCP 61 -- eventually some agreement about what our security goals are. - Russ Housley: We don't have "running code" -- a culture of privacy awareness, are we asking the right questions, etc. Once we get that far, we'll be in a better place to write BCPs
Nick Doty: Regarding draft-hansen, we use terminology for goals and properties they wish to achieve and avoid, but they don't use the full slate of terminology in the draft.

- Reference: http://tools.ietf.org/html/draft-hansen-privacy-terminology

Alissa Cooper / CDT / Privacy Consideration for IETF Documents
There appears to be a latent privacy concern within the IETF, and it's just a matter of surfacing it.
Bernard Aboba: For people who have worked on privacy issues in the IETF, would the questions proposed in this presentation be useful/would they have been useful when you were working on privacy-related standards?
Cullen Jennings: I worked on SIP (3323 and 3325) in the IETF and our privacy solution was never implemented by vendors.
Jon Peterson: RFC 3323 asked these kinds of questions, RFC 3325 didn't
Nick Doty / UC Berkeley: how were these questions selected instead of those from the older draft-morris-policy-considerations-00.txt document?

Hannes Tschofenig: draft-morris-policy-considerations-00.txt covered generic public policy aspects and did not purely focus on privacy. We removed the privacy aspects from that document and published https://datatracker.ietf.org/doc/draft-morris-policy-cons/ (for the generic public policy part) and http://tools.ietf.org/html/draft-morris-privacy-considerations (for the privacy components). [Hannes Tschofenig: I believe there was also a question about the relationship between the policy considerations and the privacy considerations document. There was the question whether the policy consideration related questions are also applicable to privacy.)
Thomas Roessler/ W3C: Another set of questions for building a model: Who needs to make a decision? Who needs to mediate? Who needs to enforce? Also, many of the privacy issues don't lie within the protocol itself, but rather with typical implementation. They must be asked.
John Klensin: IETF develops protocols and tools, not deployments. They can't assume architectures and implementations.
Lucy Lynch / ISOC: Russ noted that security considerations got better with standardized term. [Not sure what the previous sentence means.-Alissa] Also, privacy is a multistakeholder problem, including overlap between security space and various organization
Frederick Hirsh: Thomas' point about considering the system is valuable rather than just being focused on the bounds of the RFC. Also, information minimization is another useful and implicit principle that should be made explicit
Brett McDowell/Paypal: Perhaps usability and system-level consideration is the common thread between IETF, W3C and other standards-setting organizations?
Olaf Kolkman: The average engineer will need to spend four hours on the privacy considerations section. Also, like security, privacy is an expert art that is not easily obtained, and it would be useful to draw in the artists and others with expertise and interest into [This is missing the end of the sentence.-Alissa]
Susan Landau: I'm asking general questions, and I think these questions need to be asked: why are you collecting information, how long is it retained for, ... Also, this thinking should be done at the beginning, rather than at the end of the draft (even -00)
Cullen Jennings: expects this process to be a huge timesink
Sam Hartman: The IETF can't answer Susan's questions because they are out of scope of the IETF's mission and also because privacy problems can occur even when data isn't collected
John Linn: These questions may allow for us to do better than the fragmented inculcation of security into the IETF and RFCs
Christine Runnegar/ ISOC: We seem to have rough consensus that a "Privacy Considerations" section would be valuable in RFCs.
Dave Crocker: Susan's point about data retention piqued something in my mind: we need to make a distinction between privacy technologies and things that have major privacy implications and then other protocols and tools used in normal Internet operations that don't have their own privacy considerations, but could be utilized in operations to get into privacy issues
John Klensin / IAB: Metaquestions: Why do you believe that implementors will do anything to address any of the questions you're asking in privacy considerations? What makes you think they aren't lying to you?
Thomas Roessler: The W3C arguably also designs tools and protocols, although they have a rough idea about the distributed system in which the protocols will be implemented/executed. Why can't the IETF do this? How far can they think about the implications within a "protocols and tools" scope?
Jon Peterson: There is no clean line. We want to be careful that we don't become system builders but rather build protocols that can be architected and deployed in a number of ways. IETF seems to have a good track record of thinking about the implications and considerations of the protocols without thinking about the details of concrete implementations/deployments.

[@@ I seem to recall that the "systems" vs "tools" discussion was with Jon Petersen, not John Kensin. -tlr @@]

4:30pm - 5:00pm - Wrap-up for the day; plan for tomorrow
(Trent Adams)

Trent / ISOC / Wrapup
Cullen Jennings/ Cisco: Correction: We talked about economics but we never talked about turning off ads completely.
Brett McDowell/ Paypal: Observation: We never talked about getting rid of tracking altogether.
Kevin Trilli/TrustE: Observation: We spent a lot of time on control ...
Jim Fenton: Beyond smart devices, anything that doesn't have a conventional browser should be considered along with the blu-ray player issue

5:00pm - 6:00pm - Refreshment Break

6:00pm - Dinner: ROYAL EAST RESTAURANT
Location: 792 Main Street, Cambridge, MA 02139-3510
(Sponsored by ISOC, and Nokia Siemens Networks)

Thursday, December 9
===================
8:00am - 8:30am - Breakfast Coffee
8:30am - 10:15am - Cross-site Data Exchange Issues

Covered topics:
Data sharing between sites, which includes delegated authentication
(OAuth/Facebook Connect), and identity management (e.g. OpenID).

Basic structure:
1) This is how systems works now.
2) How should they work (desired state)?

Example to stimulate discussion is Facebook Connect, OAuth, etc.
(Blaine Cook).

Moderator: Thomas Roessler & Hannes Tschofenig

Blaine Cook: Yesterday's discussion was a bit esoteric. The Internet is fundamentally about sharing information. My major concern is about loss of information. I want security enhancing technology, not privacy enhancing technology. For me, the threat model is losing touch with friends and family.

Question: How is Facebook a threat to forgetting information?

Blaine Cook: It is not, Facebook is operating across various national legal regimes and so it is harder to enforce their policies.

Blaine Cook: oAuth is not a good solution; it doesn't let you securely share information with friends. (The talk is going to cover this use case.) OpenID I & II are dead because you have to use this URL to log in every where. But nobody "gets it". Systems don't have tools that dereference and interpret the Open ID URIs. The ideal solution would have the following characteristics - usable identifier, globally routable, globally unique, and free. Facebook connect is only valuable because there is only one connect. If there was a Twitter and Google analog, Facebook connect's value would be diminished.

Craig Wittenberg: A suggestion - Separate the authentication and the sharing processes.

Brett McDowell: No one has found a solution that works for the browser-based web that goes beyond the "NASCAR problem" (referring to the noise of brand logos of top-tier OpenID providers displayed as buttons, akin to the brand stickers that adorn NASCAR racecars. See http://factoryjoe.com/blog/2009/04/06/does-openid-need-to-be-hard/)

Blaine Cook: a possible solution - throw away authenticators. But users don't get the concept of using a random username, and a previously established password.

Paul Trevithick: There are several possibilities for identifiers - Globally unique identifiers, one time identifiers and opaque identifiers.

Karen Sollins: an important use case is capabilities (groups of rights which represent administrative job responsbilities). This is what we used to have.

Blaine Cook: Capabilities are an important use case.

Karen Sollins: What is the minimum amount of information one needs to make things happen? What is involved in conveying it to the user?

Paul Trevithick: Restatement - There are different use cases. To solve the discovery problem, and to address the usability issue, Blaine's solution is to use email addresses.

Thomas Roessler: How do we build a system that does not depend on a particular company or resource?

Sam Hartman: Use of email addresses is the way to go; but how do we address the throw away problem?

Also, when we talk about capabilities, we need to worry about the fact that assertions on the web are copyable - "i am over 18" can be shared among friends.

Benjamin Heitmann: Threat model is a very fuzzy term. We need to specify it more. Maybe give people an explicit choice of private vs. public information?

Dave Crocker: I'm troubled by the sweeping generalizations that have been made: users can't deal with a number of different ID's - but today we find this is the case on the web.
We need to be careful about giving people options versus burdens.
Repurposing an email address is sometimes just that and at other times it is just a format, and not necessarily a valid email address.

Olaf Kolkman: Observation - we are having an engineering session, rather than a discussion on privacy, which is what this workshop is about.

Cullen Jennings: URIs are good for some applications. But we can't always use URIs for authentication. For instance, we could have a set up where URIs are only valid over an https: connection

Blaine: capability URLs are usually used when there is a social decision to share a particular resource.

Alissa Cooper: Systems right now are designed under the assumption that the identity provider has a large amount of data on users. Is it o.k. to have large firms like Google and Facebook provide identitites?

Arvind Narayanan: There is a problem with the "possession model" that assumes that anyone who can obtain a URI has authorization to dereference the information. Capability URIs can leak everywhere, not just through intermediaries.

Balachander Krishnamurthy: It's a mistake to focus on Facebook. Only about a quarter of the Internet is on Facebook. People who use IDs and passwords to login to many sites face these threats as well. Also, it is not practical for users to have to provide consent for all types of data sharing.

Karen O'Donahue: We also have to take into account changes in user behavior. Things like deleting wall posts, deactivating accounts - we didn't observe this five years ago, but we do today. Users have a better understanding of privacy issues today.

Stefan Winter: There is support in IEEE/IETF protocols for anonymity, not so much in web standards.

Fred Carter: Are we advocating an architecture that facilitates access without taking into account the policies that dictate that access?

Blaine Cook: It is possible to create secure associations over insecure media. There is an Android app - "Redphone" (based on ZRTP) that is an example of this. We can start with something insecure (SMS) and use it to enable secure voice calls (ZRTP).
(see http://blogs.forbes.com/firewall/2010/05/25/android-app-aims-to-allow-wiretap-proof-cell-phone-calls/)

Fred Carter: Blaine is advocating layering.

Frederick Hirsh: Have heard about different interesting technologies, but not clear what our requirements are at a higher level.

Summary

Thomas Roessler: We heard about requirements that work on all the layers of the web. We need to stop relying on a handful of identity providers. What types of identifiers should we use? There are really simple protocols out there - maybe we need to build on top of that to provide privacy. But the simplicity of the protocols could mean that their ability to provide privacy is limited.

10:15am - 10:30am - Coffee break

10:30am - 12:00am - Cross-site Issues (cont.)

3) Is there a a way to get to the desired state?
(min. 45 min for item #3)

What new mechanisms are needed?

What are the implications for existing and
under development IETF/W3C protocol mechanisms

Lorrie Faith Cranor: Often privacy gets left out of the standards setting process and is left to the implementers. But in some of those cases, privacy is not implemented.

Jim Fenton: With Caller ID, it used to cost more to get the system rather than opt out and not have the service. However, over the years the dynamics of Caller ID changed and sending Caller ID became the default. Now it is necessary to pay to *avoid* sending Caller ID.

Jonathan Mayer: There is something called an "adhesion contract".
[A type of contract, a legally binding agreement between two parties to do a certain thing, in which one side has all the bargaining power and uses it to write the contract primarily to his or her advantage. See: http://legal-dictionary.thefreedictionary.com/Adhesion+Contract].

Traditionally, consumers haven't had much choice in many situations; to obtain goods and services they have to agree to the offered contract. It is "take it or leave it".

So, how do we make sure we don't get privacy policies of adhesion?

Kasey Chappelle: Today we do have privacy policies of adhesion. The question is how can we move away from that.

Lorrie Cranor: If we can condense privacy practices to a few icons, that might go a long way towards establish good privacy frameworks.

Thomas Roessler: If we have policy mechanisms, we can customize data processing across the Web.

Ari Schwartz: Without appropriate legal frameworks and enforcement mechanisms, how do we ensure that people tell the truth?

Susan Landau: All questions about user privacy are externalities. with environmental protection, the issue is that polluters can offload costs onto third parties, while reaping the benefits. We see the same thing

with privacy -- the core problem is that the technology, does not make
any aspect of the process (privacy policies, user preferences, etc.)
transparent. Do people understand things well enough in this realm?

Thomas Roessler: We have to really look at what the incentives are.
There are social incentives, there are impedance mismatches when it comes to
business processes, protocols. Maybe we are looking for a magic
recipe for privacy enhancing protocols that doesn't exist.

12:00am - 1:00pm - Lunch
1:00pm - 3:00pm - What are our conclusions? /
What is the take-home work?

Discussion of workshop administrative issues.

A mailing list will be created for workshop participants: Priv-ws@elists.isoc.org

The following materials will be made available online after the workshop:

- Final Agenda
- Presentations
- Papers (if you don't want your paper to be made public, please notify the organizers)
- Minutes (if you took notes, please send them to the organizers ASAP)

With respect to minutes, once they have been produced they will be sent to the workshop mailing list; participants will then be given an opportunity to correct them, or to request that remarks not be attributed to them.

Action items: (slides: ActionItems_workshop.pptx)

IETF Actions

- Create privacy directorate
- Investigate possible research work in privacy (IRTF)
- Explore what the IETF can learn from Tor (and vice versa)
- Support for continuing "Privacy Considerations" work

Apply it to some IETF protocols
Covers HTTP/non-HTTP-based Protocols

- Fingerprinting Considerations for IETF protocols (TCP, HTTP, SIP, etc.)

Hannes Tschofenig: There are glossary-building efforts going about privacy. Terminology for privacy needs to be described in a system perspective so people will adopt it.

John Klensin: Do we really need a huge privacy glossary?

Hannes Tschofenig: Probably not. But we do need to suggest adoption of a few terms.

Sean Turner: Some potential followup items:

Sean Turner: We will create a Privacy Directorate, to enable review of privacy aspects of protocol designs. Gaining experience in reviewing documents for privacy aspects will provide experience that could be useful in developing documents such as draft-morris-privacy-considerations (see http://tools.ietf.org/html/draft-morris-privacy-considerations). One document which might be useful to review for privacy considerations is "IP Flow Anonymization support" (see http://tools.ietf.org/html/draft-ietf-ipfix-anon).

John Klensin: We need to be careful about creating yet more bureaucracy around review and approval of documents. Are we going to require a privacy considerations section in every document? Is their IETF consensus for that?

Bernard Aboba: There might well be IETF consensus -- *against* :)

Sean Turner: Initial focus will be on documents that relate specifically to privacy aspects.

Nick Mathewson: Tor isn't yet ready for standardization. We have quite a bit of work to do.
Sean Turner: Perhaps some aspects could be looked at, such as onion routing?
Sean Turner: How do we drive the ‚Äúdo not track‚Äù mechanism?
Jonathan Mayer: The technical direction of ‚Äúdo not track‚Äù has not been decided yet. It could involve a Do-Not-Track header, or it might involve another mechanism.
Sean Turner: Agreed. Would it make sense to set up an IETF mailing-list to get the discussion going?
Jonathan Mayer: We need some possible experimental activity.

W3C Action Items

- Formation of W3C Privacy Interest Group
- Fingerprinting of W3C protocols
- How to help
- API Minimalization
- Referrer header & privacy
- Privacy Considerations for Non-Browser-based Protocols
- Usability:
  - Whitepapers & Best Current Practices (BCPs)
  - Usability Considerations in W3C specifications

Action items for Both/Miscellaneous

- Debate "Do not track" technical mechanisms and policy consideration
- Technical analysis
- Possible experimental activity
- Cookie / cross-site management
- Terminology
- Threat Models (based on system-level view)
- Use Cases
- Explain why we care about correlation/linkability (examples in the workshop report)
- Investigate policy expression languages

Question: What are the technical topics that W3C is focusing on?
Nick Mathewson: The outcome of the workshop should give guidance to program agent developer about how to treat data sharing/data information.

Question: Should privacy usability be on the list of W3C actions?
Thomas Roessler: There is a fine line between specifying useful/important directions and implementation for usability
Question: Should the cookie policy mess be on the list of W3C actions?
Thomas Roessler: we need a common terminology about privacy of the Web, suggested by the W3C.

Frederick Hirsch: We need apriori threat models, considering privacy from a system perspective.
Karen Sollins: We need to examine use cases, starting from a legal/business perspective, then proceeding to a discussion of technology.
Susan Landau: The IETF needs to be able to consider privacy aspects of protocols under consideration, such as the privacy tradeoffs considered within protocol design.
Nick Doty: We need to improve our ability to think about privacy threat models. Threats exist from attackers, from mediators and from users. These threats have different aspects depending on where the threat is coming from. It is necessary to make assumptions about the threats originating from each role, and to correctly capture that within the threat model. Today threat models are often incomplete (e.g. a 'private browsing' threat model that only captures threats from other users), and if some parts are ommitted or missed, then threat modelling will be useless. This is critical since it is the threat model that enables the translation of privacy concepts into engineering terms.
Dave Crocker: The body of privacy work seems to be concerned with , "data & attributes". This relates to the static aspect of "data at rest". The IETF is concerned with a different aspect, which is the protocol dynamics and the degree of privacy afforded by the protocol design.

Lorrie Cranor: Is there any role for standards organizations (IETF/W3C) in driving the policy for privacy? For example, should they work on how to express privacy policies?

Alissa Cooper: The IETF has worked on privacy policies in the past (GEOPRIV WG).

Sam Hartman: It might be useful to put effort into recalling what information was disclosed after-the-fact.
Susan Landau: What about work on data retention policies?
Jim Fenton: It's important to understand that not all uses of the HTTP protocol involve browsers. So privacy proposals need to include consideration of non-browser users. Do the proposals work for those users?
Sam Hartman: We did not cover privacy at the network layer beyond Tor. We also did not cover the aspects of privacy beyond web servers, such as the privacy aspects of backends (e.g. databases, etc.).
Eric Rescorla: Best practices for operator data retension could be useful. However, there might be better organizations than the W3C/IETF to address those issues.
Susan Landau: There are data for identifiers within the protocols. We need to distinguish ephemeral data from other
types of information which travel over the network. This should be specified within the protocol.
Marc Linsner: If we could define what bits of data mean from protocol perspective, there will be no ambiguity on who should be responsible.
Nick Mathewson: to make some stuff work in http, you need to retain cookies. Some implementations will discard the information quickly, others will not. We need guidelines for retaining or processing a certain type of data.
Lucy Lynch: We need to define how to use different types of identifiers in different contexts. The identifier is created with a scope under certain context, and the usage needs to be appropriate, given that.
Frederick Hirsch: We need to better understand the unintended consequences of choosing identifiers, the secondary uses.