Approaching Privacy over ICN: Values-in-Design Approach

Jeff Burke, Katie Shilton, Nicholas Proferes, Dustin O'Hara

ICNRG Interim Meeting, Buenos Aires
April 3, 2016

Purpose

1. Suggest starting points for the ongoing discussion of privacy support in ICN.

2. Briefly introduce a perspective on privacy that comes from socio-technical studies of privacy. Distinguish *ethical / social motivations* from *architectural design goals* from *mechanisms*.

3. Dig into some reasons why the “TLS everywhere” concept should not be the sole point of departure for ICN design.

4. Suggest a design space to be explored.

First, some clarifications

Critical threads of the TLS-baseline position

- Application model is two-party conversation between individuals and services [centrally administered, broadly distributed, with sufficient resources to be provided at global scale.]

- Important property of forward secrecy (via ephemeral keys) of the data exchanged during the session.

- Protection of end-user request confidentiality.

Critical threads of the TLS-baseline position

- Application model is two-party conversation between individuals and services [centrally administered, broadly distributed, with sufficient resources to be provided at global scale.]

  => What about other current and future applications models?

  => What about when these assumptions harm privacy?

- Important property of forward secrecy (via ephemeral keys) of the data exchanged during the session.

  => Key granularity and lifetime control not unique to TLS (see NAC).

  => How long is that Google Drive TLS session connected?

- Protection of end-user request confidentiality.

  => Can we achieve this on its own?

Critical misinterpretations of the NDN approach

- Everything is in the clear.
- All keys are long-lived and coarse-grained.
- Assume all data around forever.*
- Socio-technical implications of the work are not considered.

See tech reports NDN-0034, -0030, -0036 and more recent application designs (forthcoming), as well as Shilton, K., J. Burke, k. claffy, and L. Zhang. "Anticipating Policy and Social Implications of Named Data Networking," to appear in Communications of the ACM, 2016.

More accurate characterizations

- Multiparty information dissemination without reliance on (but also without excluding) centralized services is an important motivation.
- Synchronization of collections rather than conversational sessions are the primary high-level transport model.
- Both intentional and opportunistic communication is potentially common.
- Cleartext names are powerful tools for applications. (But *to whom* are they clear?)

With that in mind…

Some proposed reframing

Given the R in ICNRG, perhaps iterate on the following:

- Distinguish *ethical / social* motivations from *architectural design goals* from *mechanisms*.
- Articulate the motivations leading to TLS everywhere and *other* critical motivations.
- Consider what the architecture does (or can do) holistically to address those motivations.
- Turn these considerations into proposed design approaches based on existing security mechanisms and new architectural assumptions.
- Explore tussle between best practices in a network for point-to-point communication vs. an information dissemination network.
- Yields an evolving understanding of a design space that may have more than one available mechanism.

On Privacy

- Privacy is important.
- Privacy is in disarray. (Solove, 2006)
- Privacy is a spectrum.
- Key interpretations of privacy are non-technical.
- There are other values in addition to privacy.

Historical Understanding of Privacy

Protection from:

- Intrusion on the the seclusion or solitude of an individual,
- Public disclosure of private facts about an individual,
- Publicity of an individual that places them in a false light,
- Or the appropriate of an individual's likeness for someone else's advantage.

Information privacy

- Informational privacy mostly understood around the public disclosure of private facts about an individual (such as the leaking of passwords, credit card information, medical history, etc.).

- Conceived as a binary.

- This conceptualization of privacy glosses over an incredible variety of ways in which privacy is a function of situational context.

Privacy as Contextual Integrity

- Nissenbaum (2004) argues for conceptualizing privacy as about contextual integrity: There is a context for the flow of information, and violations to this context are what cause privacy concerns.

- The three typical principles of concern:
  1. limiting surveillance of citizens and use of information about them by agents of government,
  2. restricting access to sensitive, personal, or private information, and
  3. curtailing intrusions into places deemed private or personal.

- How does TLS to Facebook, Google, Dropbox, etc. address #1-#3?

TLS

- Two-party connection assumption – so all multi-party communication is service-mediated. (*Is this good for privacy?*)

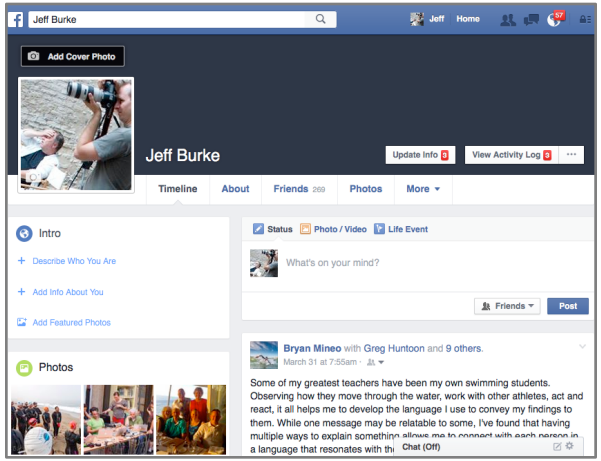
- Within that communication model, in ICN terms:
  - Provides request confidentiality (what is consumer asking for?)
  - Provides publication confidentiality (what is producer providing?)
  - Provides publisher data integrity / provenance (who is producer?)
  - What about consumer identity?
  - Still have endpoint addresses and perhaps DNS lookups. (Service locator in ICN case?)

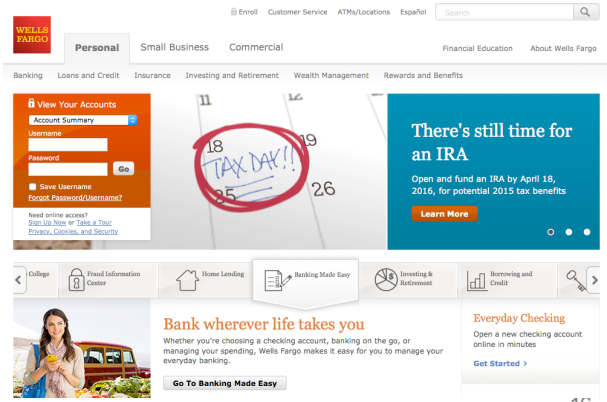
Can we separate?

1. Data integrity
2. Publisher confidentiality
3. Consumer request anonymity
4. Two-party conversational model

Use case #1: Global service in current Internet model

- Highly successful.
- Centralized control.
- Privacy of the TLS session != Privacy of my data.
- What type of business model relative to my data?





What if the best thing we could do for privacy was to enable the application itself to be reformulated?

With respect to privacy, my relationship to my bank is not the same as my relationship with Facebook.

Hypothetical:

- ICN-based decentralized social media
- Publish-anywhere, service-as-rendezvous
- User control over collection and use of data
- Opt in for data analysis for algorithmic curation
- "freemuim" business model, opt in to viewing adds, micro-payment to users for viewing ads

image credit: ibtimes.com protesting NSA data collection

Use case #2: Public data in an ICN model

- "Piles of digital information and the algorithms to analyse them tend to be good for those in power."
- Y. Benkler (Harvard) now sees data as a force for recentralisation that allows "the accumulation of power by a relatively small set of influential state and non-state actors".
- Transparency of public data while protecting request confidentiality for open data?

The Economist

Special report:
Technology and politics

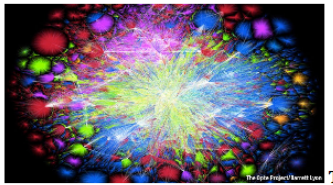
Living with technology

The data republic

To safeguard democracy, the use of data should be made as transparent as possible

Mar 26th 2016 | From the print edition

"TECHNOLOGY IS NEITHER good nor bad; nor is it neutral," said the late Melvin Kranzberg, one of the most influential historians of machinery. The same is true for the internet and the use of data in politics: it is neither a blessing, nor is it evil, yet it has an effect. But which effect? And what, if anything, needs to be done about it?



The internet, but not as you know it

Jürgen Habermas, the German philosopher who thought up the concept of the "public sphere", has always been in two minds about the internet. Digital communication, he wrote a few years ago, has unequivocal democratic merits only in authoritarian countries, where it undermines the government's information monopoly. Yet in liberal regimes, online media, with their millions of forums for debate on a vast range of topics, could lead to a "fragmentation of the public" and a "liquefaction of politics", which would be harmful to democracy.

The ups and downs of the presidential campaign in America and the political turbulences elsewhere seem to support Mr Habermas's view. Indeed, it is tempting to ask whether all this online activism is not wasted political energy that could be

What if a core contribution of ICN to a free and open society was to decouple protections for request anonymity from publisher confidentiality and control?

Example of National Archives, Data.gov, etc.

- mandate to make records and collections available to the public
- archivists publish history collections online
- machine readable transcriptions, metadata, and audio files
- Oral history player on archives website and API for developers to pull from collection
- *Why should this data be encrypted, if we can provide request confidentiality?*
- *Is widespread, distributed storage and dissemination of public information a social goal?*

Use case #3: **Connecting the Next Billion(s)**

Rural village with limited bandwidth

- wireless mesh network
- gateway node connecting to broader internet locally hosted services
- limited electricity = nodes going on and off
- 50% of communications are to endpoints local to the mesh
- inbound traffic is often similar or duplicate content
- ICN can help with intermittent connectivity and lower costs of upstream bandwidth. Doesn't rule out dynamic data / interaction with services; provides more bandwidth for it.
- Conflating request anonymity with publisher confidentiality here (e.g., encrypted YouTube) hurts us here.

- *Will services really be co-located at these edges in the foreseeable future? Would they help or hinder privacy? Can ICN help keep local data exchange local?*

image credit: village telco

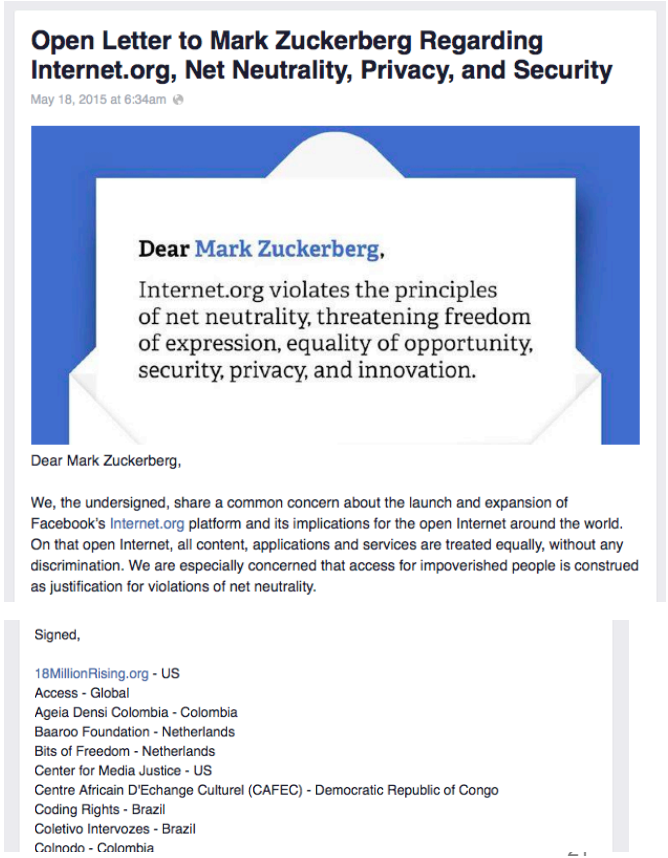
Can ICN support the privacy and agency of the next billions, by enabling information exchange more effectively than a "big services" mentality?

Excerpts of the letter:

Net neutrality: […] We urge Facebook to assert its support for a true definition of net neutrality in which all applications and services are treated equally and without discrimination — especially in the majority world, where the next three billion Internet users are coming online — and to address the significant privacy and security flaws inherent in the current iteration of Internet.org.

Privacy We are very concerned about the privacy implications of Internet.org. Facebook's privacy policy does not provide adequate protections for new Internet users, some of whom may not understand how their data will be used, or may not be able to properly give consent for certain practices. Given the lack of statements to the contrary, it is likely Internet.org collects user data via apps and services…

Security: The current implementation of Internet.org threatens the security of users. The May 4 update to the program prohibits the use of TLS (Transport Layer Security), Secure Socket Layer (SSL) or HTTPS encryption by participating services. This inherently puts users at risk, because their web traffic will be vulnerable to malicious attacks and government eavesdropping.



Open Letter to Mark Zuckerberg Regarding Internet.org, Net Neutrality, Privacy, and Security

May 18, 2015 at 6:34am

Dear Mark Zuckerberg,

Internet.org violates the principles of net neutrality, threatening freedom of expression, equality of opportunity, security, privacy, and innovation.

Dear Mark Zuckerberg,

We, the undersigned, share a common concern about the launch and expansion of Facebook's Internet.org platform and its implications for the open Internet around the world. On that open Internet, all content, applications and services are treated equally, without any discrimination. We are especially concerned that access for impoverished people is construed as justification for violations of net neutrality.

Signed,

18MillionRising.org - US
Access - Global
Ageia Densi Colombia - Colombia
Baaroo Foundation - Netherlands
Bits of Freedom - Netherlands
Center for Media Justice - US
Centre Africain D'Echange Culturel (CAFEC) - Democratic Republic of Congo
Coding Rights - Brazil
Coletivo Intervozes - Brazil
Colnodo - Colombia

Use case #4: **Protecting unpopular content**

- Book sellers in sell banned political books; they are kidnapped by authorities
- Protesters begin publishing machine readable versions of banned books
- Protesters create banned reading list hotspots using inexpensive devices that are hidden in public spaces broadcasting WiFi and bluetooth signals that provided access to the banned books.
- *Can this type of local publishing be best supported by TLS sessions, or ICN-style opportunistic dissemination?*

image credit: theatlantic.com

Use case #5: **Protecting highly personal content**

Los Angeles Health Department Database

- 2nd largest health system in the US, services over 670,000 unique patients
- builds digital medical records systems, available to all department sites
- community centered approach to chronic care, working with patient's immediate social network of friends and family to help promote wellbeing and ongoing care
- social and technical protocol for how and when authority is managed by others

Who should control the keys that encrypt individual health data?
Should ownership and control (agency) of our data be a goal?
Or is centralized control the most robust privacy option?

Opportunities

- Employ use cases and broader privacy challenges to **illuminate a design space** that includes not only TLS-like sessions but other communication models as well.
- If protecting privacy is a critical social goal or principle, consider where ICN-based models could have **more holistic privacy benefit** than simply providing secure point-to-point connections.
- **Examine where the "distributed service" model fails** – for example, where personal agency, privacy, and/or innovation emerge from local communication capacity or other situations well-supported by ICN.
- Recover open data. Explore specific mechanisms to **provide request confidentiality** without requiring content encryption.
- Continue to explore **secure multi-party information dissemination** over ICN that is less infrastructure-reliant and meets forward secrecy requirements with desired granularity.

Thank you!

jburke@ucla.edu