Session 1: Web Next - decentralized, privacy-preserving, and resilient

Transcribed by: M. Taimoor Tariq, Ammar Tahir (UIUC)

Global Content Revocation on the Internet: A Case Study in Technology Ecosystem Transformation

Narek Galstyan *(UC Berkeley and ICSI); James McCauley (Mount Holyoke College); Hany Farid, Sylvia Ratnasamy (UC Berkeley); Scott Shenker (ICSI and UC Berkeley)*

Summary:

- Leaked media on the internet can be problematic. But once data is on the internet, there is no ownership and no way to track it down and remove it from all websites. There is a need for a revocation system. Internet Revocation System allows users to revoke the data they owned from the Internet. The idea is to maintain a signed ledger. Paper also proposes TET to bootstrap and motivate big platforms to take up the system.
  Key Idea: Design an internet revocation system by maintaining a globally anonymized ledger where ownership can be tracked

Questions:

1. **How do you define who owns the content? You can change media by changing a few pixels/bits.**

   Not fully solved for now. Start with a general idea as a starting point and then work on edited photos. There should be a seperate review process to see whether the picture is copied/edited.

2. **What is the threat model? Can users be malicious?**
   For bootstrapping, it relies on the good will of users.

The Decoupling Principle: A Practical Privacy Framework

Paul Schmitt *(University of Hawaii/INVISV); Jana Iyengar (Fastly); Christopher Wood (Cloudflare); Barath Raghavan (USC/INVISV)*

Summary:

The decoupling principle essentially means that third parties should only know one of 2 sensitive info pieces:

- Sensitive data
- Identity

If a third party has a view of both, that can be problematic. Decouple the 2 and ensure that any third party can only have access to one at any given time.

Questions:

1. **Are you able to reason quantitatively about the decoupling system you are connecting to? How to enforce decoupling to avoid collisions.**

   We haven't explored quantitatively analyzing decoupling. Degree of privacy could be one metric. Enforcement: Difficult, becomes a policy problem where a lot of legal problems start to kick in.

Reflections on trusting distributed trust
***Emma Dauterman, Vivian Fang**, Natacha Crooks, Raluca Popa (UC Berkeley)*

Summary:
Problem: Secret key stored locally, but if local copy lost key is gone. Can't simply store it online because the third party has it.
Solution: Distributing trust (e.g splitting and storing your keys) across different trust domains can help backup private data. If one trust domain is compromised, the attacker still cannot recover data e.g. fragmented secret key. Developers can then become the central point of attack. Auditing the code and deployment can keep developers in check. Hardware can also become a central point of attack, diversifying across TEEs.

Questions:

1. **How much is it going to cost because of having to deploy across a large number of trust domains?**
   You will have to pay extra to ensure the high level of security you can get (*More trust domains == Better privacy?*).

2. **How do you authenticate when a key is split across domains?**
   Use multi-party compute protocol but it will require authentication of users.

Making Links on Your Web Pages Last Longer Than You
***Ayush Goel**, Jingyuan Zhu, Harsha V. Madhyastha (University of Michigan)*

Summary:
With time, hyperlinks either redirect to a 404 page (link rot) or content changes too much (content drift). Page snapshot (status quo) makes a copy of the page and stores it which is static. Snapshot should be the last resort, not the very first one. Importantly, the purpose of the link should be taken as context to determine what kind of snapshot should be shown. E.g. in some cases you want to link to dynamic content so snapshotting wont work, but then in some cases you want to link to static content, in that case direct linkage wont work. So how to come up with a uniform, backwards compatible solution.

Questions:

1. **Is duralink supposed to be a scalable solutions**
   Run as a third party. Merge with CDNs. As a provider, i will register all my links with duralinks and will allow users to access these pages to avoid link rot

2. **We may need different semantics for different links?**

   Links need to capture the intent of the user: annotation based techniques, author annotates link based on author perceived context about intent.

End of Session Discussion

How important are conflicting policies about content revocation? E.g. policy of the US government vs Twitter?
Different jurisdictions have different rules, this is a first step to get attention from the government and big players.

Auditing as a way to solve problems. Is it a good idea?
There is value to it and it simplifies and makes the problem easier as well by outsourcing it. Auditing also provides some level of guarantee that problems is being improved but only helps, does completely solve

How to prove ownership? The person who takes the picture, or the person who is in the picture.
Only consider the photos where there is a clear owner. The one who takes the photo is considered the owner in this model.

How to do the images solution at scale, e.g. at insta, fb if you want to review the photos, how to do it at scale.
Most of this image checking does not need manual review, it can be automated, e.g. through metadata or watermarks.

How to convince people to care about security?
We can build systems which are better for them but don't know if you can motivate people to care about security. It's important to maintain general purpose, easy to use tools so we can lower the barrier for people to understand security and care about security.

Web 200 years from now. Will it be a graveyard of media?
For starters, we don't even know about whether data will be there for 200 years. Preserving it will be hard in a scalable manner, e.g. in system storage. Transcribing it on glass, MSR Cambridge projects Silica to store data in an immutable way for 1000s of years.

Revocation as a sign of censorship?

Since your device will be the one singing the image, therefore you will have ownership and you will have the right to choose whether it's censored.

Hard to ensure trust in digital data creation when we are allowing users more and more rights to edit/update the content they create. But hard to argue against it as well, since if the users own the information, user also reserves the right to forget that information, invalidate those links