Lecture 2: Privacy, security, anonymity

Balachander Krishnamurthy

AT&T Labs–Research

`http://www.research.att.com/~bala/papers`

Review period

What did we discuss in Lecture 1?

- Vocabulary, history, terms
- Types of consumers
- Types of privacy
- Intro to OSNs

Recap of Lecture 1

- Right to be left alone
- Self-determination on when/how/what extent to share information
- “Informed consent”
- Online vs. Offline privacy
- Research: confidentiality/control/practice
- Origin and evolution of OSNs

Questions on last week's class?

Did anyone read papers:
[Sol07]: I've Got Nothing to Hide' and Other Misunderstandings of Privacy
[Kri10]: I know What you will do next summer
[Rot10]: A guide to privacy law
[WLM07]: Thinking about privacy, [MFSV]: Digital Footprints

Volunteers to summarize (5-10 min presentation) [Nis11, Kri10, MFSV]?

(we have at least one!)

Papers to read this week:

[Swe02] k-anonymity: a model for protecting privacy

[Nar13] What happened to the crypto dream?

[WT99] Why Johnny can't encrypt

Project Proposal—due by 2/26

Proposal of (a) your choice—it can be short or detailed or (b) my suggestion—build a personal privacy assistant

Your choice: project should demonstrate at least *several* of the following

- originality
- good choice of domain for privacy protection (FB/Twitter coo cool. Snapchat, ask.fm, kik etc. may be more intriguing).
- shows understanding of privacy loss
- demonstrates privacy protection
- examines cost for such protection (loss of fidelity, latency, etc.)
- new technologies for improving protection
- any crowdsourcing to spread or improve would be a plus

Project: PPA - personal privacy assistant

Originality here depends on depth

app that watches over your actions and tracks data dribbling

at low cost

estimate cost of privacy protection

what is the best protection/cost ratio achievable

alternate methods not implemented

can you get others to buy into it?

Coding

iOS, Android, or an extension in JS or whatever - I don't care,

just be comfortable with your choice.

Don't ask me for coding help!

In exigent circumstances TA Chris can give you general guidance.

I'll give you feedback on both the proposal and the design document due a month later 3/26.

Other students will test whatever you develop.

Onion: How to protect your personal information online

- Always log into your Gmail account in person by traveling to Mountain View, CA and letting the Google folks know its you.
- If you receive a suspicious-looking email, assiduously click on all the links and follow their instructions to learn more about the threat.
- To keep your phone data safe at all times, never unlock your iPhone screen.
- Publicly post sensitive personal information about close friends and family to draw hackers away from you.
- Hackers have been known to infiltrate public Wi-Fi networks, so make sure to switch stores or cafes every 45 seconds.
- Wear a plastic badge that says “CyberSecurity Force” to scare off snoopers
- Always be aware of your surroundings when accessing sensitive information in public. Listen closely for anyone nearby subsequently tapping on their laptop and then muttering, “I’m in”

http://www.theonion.com/articles/protect-personal-information-online,35036/

Lecture 2: Privacy, security, anonymity

- Duality of privacy and security
- Privacy and security
- The role of usability
- Crypto and its failure
- Privacy and anonymity
- Anonymization techniques
- Why hasn't privacy/security duality yielded more?

This lecture benefits from Bellovin's and Maritza Johnson's lecture notes

Duality of privacy and security

- Security is about keeping unwanted traffic from entering one's network
- Privacy is about keeping wanted information from leaving one's network
- Is this always true?
- Is it too simplistic?
- Does it yield any valuable insights for the privacy conundrum?
- What have been the success stories of security?
- How applicable are lessons from security to privacy?

Lecture 2: Privacy, security, anonymity

- Duality of privacy and security
- **Privacy and security**
- The role of usability
- Crypto and its failure
- Privacy and anonymity
- Anonymization techniques
- Why hasn't privacy/security duality yielded more?

Privacy vs. Security

- Significant variations in privacy requirements across users
- As we saw earlier, there is a spectrum: libertarians to paranoid
- Security market is different
- There are not many dissenting voices on need for security as it is viewed as a more serious threat
- Security attacks lead to disconnection from the Internet, file and other damage, loss of tangible economic assets
- Privacy damage is not viewed in the same way for reasons that are not always clear

Security

- Well studied and well understood field
- Adversary model is much better known
- Attackers have been classified into clear categories
- Attack vectors are broadly understood
- Defenses have been built over the decades
- Terms like firewall are known to many lay persons (even if they do not fully understand what it does)

Online security

- Firewalls
- Anti-virus software
- Case-hardened OS
- Secure kernel

Security as confidentiality, integrity, availability

- Who gets to access data/information — confidentiality
- Has the data been modified or destroyed – integrity (more to do with $$)
- System made inaccessible due to attacks (e.g. DoS) – availability

Security products

- Numerous commercial products available
- Most require minimal end-user configuration
- Why is this important?
- As field matures, it becomes second nature and more users can benefit
- Usability is key (something to remember for privacy)
- Hundreds of millions use pre-configured firewalls in home routers
- Many modern OS come with anti-virus systems
- Most browsers are equipped with various security alerting mechanisms – although usability problem persists

So why go into so much detail on security?

- A lot of security solutions have worked
- Can they be adapted to privacy?
- What are the challenges?

These are issues that you can use in your project

Lecture: Privacy, security, anonymity

- Duality of privacy and security
- Privacy and security
- **The role of usability**
- Crypto and its failure
- Privacy and anonymity
- Anonymization techniques
- Why hasn't privacy/security duality yielded more?

Usability

- Extent to which a product can be used...
- ... by specified users to achieve specified goals...
- ... with effectiveness, efficiency and satisfaction..
- ... in a specified context of use

ISO 9241-11: Ergonomics of HCI (1998):

Usability and security online

- How many here use encrypted email?
- Why? Why not?
- Even if your exchanges are pedestrian, are there some messages that you send/receive encrypted?
- Difficulty of managing files, let alone firewalls
- Passwords are well-studied and are still a major problem

Lecture: Privacy, security, anonymity

- Duality of privacy and security
- Privacy and security
- The role of usability
- **Crypto and its failure**
- Privacy and anonymity
- Anonymization techniques
- Why hasn't privacy/security duality yielded more?

Crypto: a brief intro

- Conversion of *plaintext* into *ciphertext* (and back) via a *key*
- A good cryptosystem won't allow for enumeration of all possible keys or get plaintext back without the key — security entirely depends on key
- Should survive brute-force attacks
- Public key cryptography allows parties to exchange data without having to make prior arrangements with different keys for encryption/decryption
- Encryption key can be public! (can receive encrypted messages from all)
- RSA is best known public key system that relies on difficulty of factorization of product of two large prime numbers

Success and failures of crypto

- Crypto to protect financial transactions works (your credit card is typically not sent over as plain text)
- Both you and merchant benefit from such encryption
- Crypto's original goals were to prevent fraud, thievery...
- ...but also to prevent snooping (i.e., to improve privacy)
- Seminal work by Chaum [Cha85]
- Security without Identification: Transaction Systems to Make Big Brother Obsolete

Crypto's difficulties

When incentives are not aligned or when usability becomes a problem, crypto is less useful for privacy

- No shared vocabulary (not easy to convince users)
- Talking about math makes even your eyes glaze over (mego)
- "Why Johnny Cant Encrypt: A Usability Evaluation of PGP 5.0" [WT99]
- PGP was a clever idea to let any pair of users exchange secure email
- GUI of PGP was targeted to make crypto accessible to naive users
- User study [WT99] showed critical failures in several aspects of usability

Privacy vs. security: so where are we?

- Many privacy issues are *worse* than security
- For example, default settings for security are often right
- In contrast, default for privacy is most permissive. Why? Incentives!
- Who benefits from more permissive privacy settings? Collectors of data
- Defaults are rarely changed
- Unlike even CAs on browsers, privacy software has to be found and installed
- Maybe no one believes there is a market yet for privacy
- Or current lack of agreement on a clear threat model.

Security's contribution to privacy

- Various cryptographic approaches to preserving privacy.
- Even recent interesting work on protecting privacy while retrieving information
- Break user's queries into subqueries to reduce risk of reverse engineering of user's intent [Yek10]
- 2010 blackhat convention: various privacy issues ranging from tracking RFID tags from significant distance, to deployment status of anonymous darknets and attacks on device privacy
- http://www.blackhat.com/html/bh-us-10/bh-us-10-home.html
- Tor: widely deployed, well-studied, with potential for use in privacy protection
- ...but still not remotely used enough for privacy

Lecture 2: Privacy, security, anonymity

- Duality of privacy and security
- Privacy and security
- The role of usability
- Crypto and its failure
- **Privacy and anonymity**
- Anonymization techniques
- Why hasn't privacy/security duality yielded more?

Privacy and Anonymity

A key component of privacy is the notion of anonymity

- Identity: Strong and weak identifiers, identifiable attributes
- Degrees of anonymity

Anonymity

- What is anonymity? Closely tied to identity
- Anonymity can thus be defined as ”without attribution to an individual”
- “On the Internet nobody knows you are a dog”
- Privacy definition: “one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity” (EU directive) http://www.dataprotection.ie/docs/EU-Directive-95-46-EC–Chapter-1/92.htm
- Need to suppress some or all of identifiable attributes

Identifiers: strong, weak, local global

- What does it mean to say that an individual may have many identities?
- Weak ids can refer to many (John Smith as name)..
- ..but many weak ids can be combined to uniquely id someone
- Local and global identifiers
- Physical attributes are a bit harder to hide (height, weight, race)
- Online attributes are much easier to alter
- Linkable attributes: dob, gender, zip (we will return to linkability)

When do we desire anonymity?

- We all seek to be anonymous at certain times in certain contexts
- We may want to search online without revealing who we are
- In the physical world it is easy (just go to nypl)
- Even with cctv outside, hard to track searches
- Possible online via tor, proxies, and other obfuscation techniques
- Visiting certain websites may risk leakage of identity
- Especially when there is a risk of false positives (searching on behalf of others)
- Comments/postings online
- When/where else to we seek anonymity?

Privacy and Anonymity

- So what is the difference?
- Anonymity protects identity
- Privacy is a deeper concept
- Even if you know who I am you may not know much about me

Lecture 2: Privacy, security, anonymity

- Duality of privacy and security
- Privacy and security
- The role of usability
- Crypto and its failure
- Privacy and anonymity
- **Anonymization techniques**
- Why hasn't privacy/security duality yielded more?

Anonymization techniques

- Context-dependent: may require significantly different techniques depending on nature of data being anonymized
- Ways of evaluating success of anonymization
- Metrics involve identifying a balance between making data safe to be shared yet have high utility

Anonymization categorization per-protocol/application

Protocol	Identity related	Personal sensitive	Organization specific	Business sensitive
BGP	Prefix	N/A	Addresses prefixes	Routing
DNS	Machine name	N/A	N/A	Specific domains, CDN use
Web	URLs, cookies, email address	Password, search strings, credit card #	Sites visited client IDs	Traffic level, commerce
P2P	IP address	Nature of content	Legality	Traffic volume
Games	IP address, name	N/A	Extent of participation	N/A

Anonymization process: How

Technique	Advantages	Disadvantages
Online	Real-time	Resource heavy (memory, CPU), may be irreversible
Off-line	Tailored to data	Not real-time, storage costs
Parallel	Multiple datasets	Inconsistent mapping of same value

Anonymization techniques: transformation

Technique	Example use	Common appl./protocol
Lossless transformation	2-way hash function	?
Semi-lossy transformation	String portions	Aggregated Web, IP, data
Lossy transformation	Map strings to numbers	All

Anonymizing personal data

- Name, address, phone numbers, CCN, SSN
- External databases can be used to partially re-identify “anonymized” strings
- Identifying components exist (area codes, bank names)
- Even non-anonymized strings are problematic [AG09]
- First 3 digits: Area number, next 2: Group number, last 4: Serial number
- AN: mailing address in SSN application form (state population dependent), GN and SN: serially allocated (can use age information)
- No randomness incorporated!

[AG09] Predicting Social Security Numbers from Public Data
Acquisti and Gross, Proceedings of the NAS, 2009

Lecture 2: Privacy, security, anonymity

- Duality of privacy and security
- Privacy and security
- The role of usability
- Crypto and its failure
- Privacy and anonymity
- Anonymization techniques
- **Why hasn't privacy/security duality yielded more?**

Why hasn't privacy/security duality yielded more?

- Security research has pointed out potential privacy leakages
- However, in terms of protection there has been limited reuse
- Structural differences: motivation of adversaries, nature of what is being protected
- Usability is presumably a key difference
- Most security solutions come pre-installed with little or no direct end-user participation
- URLs have https: in them, routers have firewalls pre-configured
- Privacy protection is at best an afterthought

Security vs. Privacy economics

- Late in the course we will discuss privacy economics in depth
- Emerging area, limited data and thus no consensus
- Cost of security failures well known and quantifiable
- Typically security impacts enterprises (cf. Willie Sutton)
- Corporations do not want to lose money
- Economic impact on individuals due to security attacks minimal: transient loss of service, exploitation as a bot etc.
- Individuals have not yet figured out value of personal data
- Corporations may be able to manage with just aggregated data

Next week: Technology–1

- Terminology and key players
- Tracking
- Technologies for tracking
- Identifying leakage
- Role of JavaScript
- Role of protocols

References

[AG09] Alessandro Acquisti and Ralph Gross. Predicting social security numbers from public data. *Proceedings of the National Academy of Science*, July 2009.

[Cha85] D. Chaum. Security without identification: Transaction systems to make big brother obsolete. *Communications of the ACM*, 28(10):1030–1044, 1985.

[Kri10] Balachander Krishnamurthy. I know what you will do next summer. *ACM SIGCOMM CCR*, 40(5), 2010. `http://www.research.att.com/~bala/papers/ccr10-priv.pdf`.

[MFSV] Mary Madden, Susannah Fox, Aaron Smith, and Jessica Vitak . Digital footprints. `http://www.pewinternet.org/Reports/2007/Digital-Footprints/1-Summary-of-Findings.aspx`.

[Nar13] Arvind Narayanan. What happened to the crypto dream? parts 1 and 2. *IEEE Security and Privacy*, March-April, May-June 2013.

[Nis11] Helen Nissenbaum. A contextual approach to privacy online. *Daedalus the Journal of the American Academy of Arts & Sciences*, 140(4):32–48, Fall 2011. http://www.amacad.org/publications/daedalus/11_fall_nissenbaum.pdf.

[Rot10] Robert L. Rothman. A guide to privacy law, 2010. http://www.privassoc.com/Documents/ICLE%203rd%20Annual%20Information%20Technology%20Law%20Seminar%209-22-2010.pdf.

[Sol07] Daniel J. Solove. "'i've got nothing to hide' and other misunderstandings of privacy". *San Diego Law Review, Vol. 44, p. 745*, 2007. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=998565.

[Swe02] Latanya Sweeney. k-anonymity: a model for protecting privacy.

International Journal of Uncertain. Fuzziness and Knowledge-Based Systems, 10(5):557–570, 2002.

[WLM07] James Waldo, Herbert S. Lin, and Lynette I. Millett. Thinking about privacy. In *Engaging privacy and information technology in a digital age*. National Academic Press, 2007. `http://www.nap.edu/openbook.php?record\_id=11896`.

[WT99] Alma Whitten and J.D. Tygar. Why johnny can't encrypt: A usability case study of pgp 5.0. In *8th USENIX Security Symposium*, August 1999.

[Yek10] Sergey Yekhanin. Private information retrieval. *Communications of the ACM*, 53(4):68–73, 2010.