Practical Privacy

Prof. Clifford Neuman
What is Privacy?

- Privacy is about Personally Identifiable Information
  - I may use the term sensitive information in discussions
- It is primarily a policy issue
  - Policy as a system issue
    - Specifying what the system should allow
  - Policy as in public policy
    - What should we require in terms of privacy
- Privacy and Theory
  - How to precisely define Privacy, how much PII is present in the information we share?
- Privacy in Practice
  - Protecting the information we don’t share.
Privacy and Security

• Privacy is a Security Issue
  – Security is needed to implement the policy
  – Compromise of the security of sensitive information compromises privacy.

• Do we trade privacy for Security
  – Franklin – “Those who would give up essential Liberty, to purchase a little temporary Safety, deserve neither.”
  – Security both depends on privacy, and can be improved by doing away with privacy.
  – The issue is private from whom
    • The policy issue again, but we know that our lack of security today stems from the inability to technically solve this policy problem.
Security v. Privacy

• Sometimes conflicting
  – Many security technologies depend on identification.
  – Many approaches to privacy depend on hiding one's identity.

• Sometime supportive
  – Privacy depends on protecting PII (personally identifiable information).
  – Poor security makes it more difficult to protect such information.
Data Breaches

- Personal Emails from Sony
- Credit card data and SSN’s
- OPM breach – especially sensitive
- Celebrity Photo Hack
Public Records

- Spokeo, etc
- Open Salary laws
- Voter records
- Implications of legislated data sharing
  - Why you might not want to give your phone number to the DMV
Case Study: Motor Vehicle Records

California Elections Code Section 2194

• (a) The voter registration card information identified in subdivision (a) of Section 6254.4 of the Government Code:
  • (1) Shall be confidential and shall not appear on any computer terminal, list, affidavit, duplicate affidavit, or other medium routinely available to the public at the county elections official’s office.
  • (3) Shall be provided with respect to any voter, subject to the provisions of Sections 2166.5, 2166.7, and 2188, to any candidate for federal, state, or local office, to any committee for or against any initiative or referendum measure for which legal publication is made, and to any person for election, scholarly, journalistic, or political purposes, or for governmental purposes, as determined by the Secretary of State.
An independent computer security researcher uncovered a database of information on 191 million voters that is exposed on the open Internet due to an incorrectly configured database, he said on Monday.

The database includes names, addresses, birth dates, party affiliations, phone numbers and emails of voters in all 50 U.S. states and Washington, researcher Chris Vickery said in a phone interview.

Vickery, a tech support specialist from Austin, Texas, said he found the information while looking for information exposed on the Web in a bid to raise awareness of data leaks.
Case Study: Conflict in Public Policy, Security Policy, and Privacy

- Third parties such as “Nationbuilder” aggregate this data for campaigns, but claim it was not their database that was found.
Could the Same Thing Happen for DMV Data?

DMV Privacy Policy

- The California Department of Motor Vehicles (DMV) is committed to promoting and protecting the privacy rights of individuals as enumerated in Article 1 of the California Constitution, the Information Practices Act of 1977, and other state and federal laws.

- DMV strives in each instance to tell people who provide personal information to DMV the purpose for which the information is collected. DMV tells persons who are asked to provide personal information about the general uses that DMV will make of that information. DMV does this at the time of collection. At the time of collection, DMV will provide information on the authority under which the request is made, the principal uses DMV makes of the information and the possible disclosures DMV is obligated to make to other government agencies and to the public.
Could the Same Thing Happened for DMV Data?

Access is Authorized:

• Request for Record Information (INF 70)
• A driver license/identification card (DL/ID) contains information obtained from an individual's DL/ID application, reportable abstracts of convictions, and reportable accidents. California Vehicle Code (CVC) Section 1808 describes this information as "public record."
• The Request for Record Information (INF 70) (PDF) form is used to request a DL/ID or VR record information on a one-time or occasional basis for:
  
  An individual's DL/ID or VR record information; other than your own. Each request is reviewed to determine that the purpose of requesting the information is for a legitimate use and that the appropriate fee has been submitted.
California has received a lot of attention in recent days for its new voter registration law, which is intended to streamline the process of signing up to vote and encourage more participation in elections. Here’s what we know — and don’t know yet — about the new law:

When people go to the DMV to obtain or renew a driver's license, or to get a state identification card, they’ll be asked for the usual information in such transactions, such as their name, date of birth and address. They’ll also be asked to affirm their eligibility to vote and will be given the choice of opting out of registering at that time. Information about anyone who does not decline registration will be electronically transmitted from the DMV to the secretary of state’s office, where citizenship will be verified and names will be added to the voter rolls.
Social Media

- Sensitive Information
  - Collection of biometric information at Facebook
- Business use of that Information
  - Sale and use of that information on behalf of other parties
- Third party use of that information
  - Facebook denies providing information to NSA
A new class action suit against Facebook alleges that the social media giant violated its users’ privacy rights to acquire the largest privately held database of facial recognition data in the world, according to a report by Courthouse News Service. Lead plaintiff Carlo Licata, represented by attorney Jay Edelson, claims that Facebook first began violating the Illinois Biometric Information Privacy act of 2008 in 2010, in a “purported attempt to make the process of tagging friends easier.”

The lawsuit, recently filed in Cook County Court, relates to Facebook’s “tag suggestions” program, which scans users’ uploaded pictures and identifies any Facebook friends they may potentially want to tag. The facial recognition technology is taken from Israeli firm Face.com, which Facebook eventually acquired. The lawsuit argues that this method of data mining directly violates users’ privacy laws, describing the facial recognition feature as a “brazen disregard for its users’ privacy rights,” through which Facebook has “secretly amassed the world’s largest privately held database of consumer biometrics data.”

The tagging feature works by scanning the faces of Facebook friends in photos and extracting facial feature data to cross-match it against their “faceprint database,” or what the company refers to as templates. However, Licata’s lawsuit alleges that Facebook “actively conceals” this information from its user base, and “doesn’t disclose its wholesale biometrics data collection practices in its privacy policies, nor does it even ask users to acknowledge them” – a practice that is illegal in Illinois.

According to the Illinois Biometrics Information Privacy Act, it is unlawful to acquire biometric data without first providing the subject with a written disclaimer that details the purpose and length of the data collection, and without the subject’s written consent. Additionally, the Federal Trade Commission also backs this same sentiment by suggesting that private companies should provide clear notice of how the technology works, what data they are collecting and for what reasons, and attain consent from the subject, before using biometric data.
Technical Means of Protection

Sensitive data must be protected in multiple places

- When it is in the hands of corporations
  - Traditional IT Security
  - Breaches in the news weekly
- On your devices
  - Encryption, access control
- On the servers you use
  - Access control for your data
Legal Protections

Privacy policies and enforcement thereof
Especially when a company is sold
Data breach disclosure requirements
EU privacy protections and flow of information
Right to be forgotten
Jurisdictional issues
Measuring Privacy

Information Theoretic approaches

Approaches based on assurance

Approaches based on assessment
Privacy not Only About Privacy

• Business Concerns
  – Disclosing Information we think of as privacy related can divulge business plans.
    ▪ Mergers
    ▪ Product plans
    ▪ Investigations

• Some “private” information is used for authentication.
  – SSN
  – Credit card numbers
Why Should you Care?

• Aren’t the only ones that need to be concerned about privacy the ones that are doing things that they shouldn’t?

• Consider the following:
  – Use of information outside original context
    ▪ Certain information may be omitted
  – Implications may be mis-represented.
  – Inference of data that is sensitive.
  – Such data is often not protected.
  – Data can be used for manipulation.
Old News - Shopper’s Suit Thrown Out
Los Angeles Times – 2/11/1999

• Shopper’s Suit Thrown Out
• By Stuart Silverstein, Staff Reporter
  February 11, 1999 in print edition C-2

• A Vons shopper’s lawsuit that raised questions about the privacy of information that
  supermarkets collect on their customers’ purchases has been thrown out of court. Los
  Angeles Superior Court Judge David Horowitz tossed out the civil suit by plaintiff
  Robert Rivera of Los Angeles, declaring that the evidence never established that Vons
  was liable for damages.

• The central issue in the case was a negligence claim Rivera made against Vons. It
  stemmed from an accident at the Lincoln Heights’ Vons in 1996 in which Rivera slipped
  on spilled yogurt and smashed his kneecap.

• Although that issue was a routine legal matter, the case drew attention because Rivera
  raised the privacy issue in the pretrial phase. Rivera claimed that he learned that Vons
  looked up computer records of alcohol purchases he made while using his club
  discount card and threatened to use the information against him at trial.

• Vons, however, denied looking up Rivera’s purchase records and the issue never came
  up in the trial, which lasted two weeks before being thrown out by the judge Tuesday.

• A Vons spokesman said the company was “gratified by the judge’s decision.” M.
  Edward Franklin, a Century City lawyer representing Rivera, said he would seek a new
  trial for his client.
Aggregation of Data

• Consider whether it is safe to release information in aggregate.
  – Such information is presumably no longer personally identifiable
  – But given partial information, it is sometimes possible to derive other information by combining it with the aggregated data.
Anonymization of Data

- Consider whether it is safe to release information that has been stripped of so-called personal identifiers.
  - Such information is presumably no longer personally identifiable
  - But is it. Consider the release of AOL search data that had been stripped of information identifying the individual performing the search.
    - What is important is not just anonymity, but likability.
    - If I can link multiple queries, I might be able to infer the identity of the person issuing the query through one query, at which point, all anonymity is lost.
Traffic Analysis

• Even when specifics of communication are hidden, the mere knowledge of communication between parties provides useful information to an adversary.
  – E.g. pending mergers or acquisitions
  – Relationships between entities
  – Created visibility of the structure of an organizations.
  – Allows some inference about your interests.
Information Useful for TA

• Lists of the web sites you visit
• Email logs
• Phone records
• Perhaps you expose the linkages through web sites like linked in.
• Consider what information remains in the clear when you design security protocols.
Protecting Data in Place

- Many compromises of privacy are due to security compromised on the machines holding private data.
  - Your personal computer or PDAs
  - Due to malware or physical device theft
- Countermeasures
  - For device theft, encryption is helpful
  - For malware, all the techniques for defending against malicious code are important.
  - Live malware has the same access to data as you do when running processes, so encryption might not be sufficient.
Linkages – The Trail We Leave

- Identifiers
  - IP Address
  - Cookies
  - Login IDs
  - MAC Address and other unique IDs
  - Document meta-data
  - Printer microdots
- Where saved
  - Log files
  - Email headers
- Persistence
  - How often does IP address change
  - How can it be mapped to user identification
Unlinking the Trail

• **Anonymizers**
  - A remote web proxy.
  - Hides originators IP address from sites that are visited.
  - Usually strips off cookies and other identifying information.

• **Limitations**
  - You are dependent on the privacy protections of the anonymizer itself.
  - All you activities are now visible at this single point of compromise.
  - Use of the anonymizer may highlight exactly those activities that you want to go unnoticed.
Protecting Data in Place

• Many compromises of privacy are due to security compromised on the machines holding private data.
  – Your personal computer or PDAs
  – Due to malware or physical device theft
• Countermeasures
  – For device theft, encryption is helpful
  – For malware, all the techniques for defending against malicious code are important.
  – Live malware has the same access to data as you do when running processes, so encryption might not be sufficient.
Privacy – Retention Policies

• PII (personally identifiable information)
  – Is like toxic waste
  – Don’t keep it if you can avoid it

• Regulations
  – Vary by Jurisdiction
  – But if you keep it, it is “discoverable”
Visibility of Addresses

- MAC or physical addresses seen only on the local network.
- IP Addresses visible to the endpoints and intermediate nodes.
- Private IP addresses behind NAT boxes (Network Address Translators) may not be visible.
- IP addresses are often transient, assigned as needed through DHCP.
  - Attributing an action based on IP address requires knowing the IP address assignment at a particular point in the past.
The Three Aspects of Security

• Confidentiality
  – Keep data out of the wrong hands

• Integrity
  – Keep data from being modified

• Availability
  – Keep the system running and reachable
  – Keeping the data available.
Security Mechanisms

• Encryption
• Checksums
• Key management
• Authentication
• Authorization
• Accounting
• Firewalls

• Virtual Private Nets
• Intrusion detection
• Intrusion response
• Development tools
• Virus Scanners
• Policy managers
• Trusted hardware
Cryptography and Security

- Cryptography underlies many fundamental security services
  - Confidentiality
  - Data integrity
  - Authentication
- It is a basic foundation of much of security.
Encryption used to scramble data
Spyware

- Infected machine collect data
  - Keystroke monitoring
  - Screen scraping
  - History of URL’s visited
  - Scans disk for credit cards and password.
- Allows remote access to data.
- Sends data to third party.
Economics of Malicious Code

- Controlled machines for sale
- “Protection” for sale
- Attack software for sale
- Stolen data for sale
- Intermediaries used to convert online balances to cash.
  - These are the pawns and the ones that are most easily caught
Economics of Adware and Spam

• Might not ship data, but just uses it
  – To pop up targeted ads
  – Spyware writer gets revenue for referring victim to merchant.
  – Might rewrite URL’s to steal commissions.
What is to be protected

• Is it the service or the data?
  – Data is protected by making it less available
  – Services are protected by making them more available (redundancy)
  – The hardest cases are when one needs both.
Why Identity is So Important

Most policy specifications are identity based
- CIA policies last week, depend on knowing who is trying to read or change data.

Most security breaches include some form of impersonation
- Malicious code runs as an authorized user
- Passwords stolen by phishing

Identifiers link data and make it findable/searchable.
- Whether right or wrong, this identification has significant impact on users.
Identification vs. Authentication

Identification

Associating an identity with an individual, process, or request

Authentication

– Verifying a claimed identity
Basis for Authentication

Ideally

Who you are

 Practically

Something you know

Something you have

Something about you

(Sometimes mistakenly called things you are)
Something you know

Password or Algorithm
  e.g. encryption key derived from password

Issues
  Someone else may learn it
    Find it, sniff it, trick you into providing it
  Other party must know how to check
  You must remember it
  How stored and checked by verifier
Something you Have

Cards

- Mag stripe (= password)
- Smart card, USB key
- Time varying password

Issues

- How to validate
- How to read (i.e. infrastructure)
Case Study – RSA SecureID

Claimed - Something You Have
Reduced to something they know

How it works:
Seed
Synchronization

Compromises:
RSA Break-in
Or man in the middle
Something about you

Biometrics

Measures some physical attribute
- Iris scan
- Fingerprint
- Picture
- Voice

Issues

How to prevent spoofing
Suited when biometric device is trusted, not suited otherwise
Multi-factor authentication

Require at least two of the classes above.

- e.g. Smart card plus PIN
- RSA SecurID plus password (AOL)
- Biometric and password

Issues

Consider attacks on the second factor and how accomplished.
Implication of Authentication Failures

Access to data (confidentiality or integrity) as if attacker were the authorized user.

For one system, or for many systems.

Failure can propagate through system.

Don’t depend on a less critical system.
How Authentication Fails

Stolen Credentials

- Passwords
- Cards / devices
- Copied biometrics

• The role of malicious code
  - GP devices can not protect credentials
Implications of Data Compromise

The biggest reason most people are concerned with data breach is:

- The data is used for authentication
- Social Security Numbers
- Credit Card Numbers
- PINs
Addressing Data Compromise

Don’t collect the data
- If you don’t need it
- Design systems so you don’t need it

Don’t use the data for authentication
- Why do we use public information for authentication:
  - Mothers maiden name
  - Password reset information
  - SSN
Identification is important for attribution

- Audit trails and logs
- Identifying wrongdoers

• Identification can be wrong
  - Attacks facilitated through compromised machines
  - IP Addresses that change
Points of Identification

- Biometric Data
- Surveillance Data
- Internet Addresses
- MAC Addresses
- Payment details
Audit and Detection

Identification data is recorded in audit logs routinely together with observed actions

– Accesses, authentication attempts, failures, etc.

Systems use tools to process this audit data and alert on suspicious actions.
Social Networks and Social Media

Services that Enable us to:

- Share our thoughts and experiences
- Record intricate details of our lives
- Create communities of like minded individuals
- Manage our relationships with others online.
Threat Vectors – Social Media

Our use of social media – dissemination
Others use of social media – retrieval
False information in social media
Reputation and permanence
Many forms of impersonation
Inferences from network analysis
Social Engineering through Social media
What we Post

Pay careful attention to what you post through social media.

We include much information we might otherwise think of as private.

We think it is going to only our friends

We think it is ephemeral

Remember what information is out there:

Fortune Teller
The Right to Be Forgotten

Article 12 DPD:

“Member States shall guarantee every data subject the right to obtain from the controller: […]
(b) as appropriate the rectification, erasure or blocking of data the processing of which does not comply with the provisions of this Directive, in particular because of the incomplete or inaccurate nature of the data;”
The Social Network

Friend graphs disclose information about social structure
  Used for investigation of criminal activities
  Used in organizational research

How The NSA Uses Social Network Analysis To Map Terrorist Networks
Social Engineering via Social Media

Assumption – Your Friends are Your Friends

Social media can be used to gather information enabling someone else to pose as your friend
  To get you to satisfy a request
  To get you to open an attachment
Malicious code will often spread through Social Media. (dorkbot)

**Cyber security sleuths warn about a new malware circulating in social media**
Ethical Issues

- Authority to search
  - Device owned by SB County
  - Court order based on showing of probably cause.
  - Genuine Probably Cause exists in this case
- Broader separate issue
  - Intentional vulnerabilities (back doors) in phone sold to other customers
  - Many problems with this
Legal Issues

- All Writs Act – a very broad law used to provide the courts authority to order.
- At issue is the burden this imposes on Apple and whether that is appropriate.
- 4th Amendment Rights not at issue in this matter as cause has been established.
- 4th Amendment is an issue in the broader discussion regarding impact on privacy of other users.
- Would complying create a precedent.
Public Policy Issues

• Impact of Required Backdoors
• Requirements to provide access to cloud data
International issues

• Level Playing Field
  – Other Countries will demand same access

• Access to cloud data across jurisdictions
  – International assistance
Tools – A Sampling

• Communication
  – Email communication
  – Website “secure email”
  – PGP / S/MIME
  – SSL / TLS
  – Virtual Private Networks

• Anonymization
  – Proxies
  – TOR

• Messaging Apps
  – Wickr

• Storage Encryption
  – Truecrypt
Email – Normal Protections

- Basic email protocol
  - Unencrypted
  - Relayed, Store and Forward
- Email can be viewed
  - On servers
  - On users device
  - In transit
Email – Better Alternatives

• For Transactional Websites
  – Financial, account management
  – Often use “secure email” hosted on site
    ▪ Certain messages accessible only by logging in to customer portal
    ▪ External emails alert users to new “secure messages”
  – Why?
    ▪ Liabilities? What is within their control?
Encrypting Electronic Mail

Mail can be encrypted in many places:

- Use of SSL/TLS to connect to gmail, yahoo, etc
- Use of SSL/TLS when downloading messages via POP/IMAP.
- Use of SSL/TLS between mail transport agents.
- Mail is in the clear on servers.

PGP provides end to end encryption and digital signatures for email messages.

- Users obtain “certificates” which certify a public key.
- Users use corresponding “private” key to decrypt messages sent to them, or to digitally sign messages they send.
- Other services provided in toolkit, to support file encryption, etc.
Email – Secure Email

- PGP or S/MIME (two different representations)
  - May encrypted message for confidentiality
  - May apply signature for integrity
  - The difficulty is:
    - Deployment
    - Key Management
      - How to know who you are encrypting a message for?
- Encryption/Privacy Debate?
  - Key escrow? Same problems as with backdoors.
Internet Communication

• SSL and TLS (HTTPS:)
  – Encrypts data sent on network
    ▪ Confidentiality and Integrity
  – Key Management for Server
    ▪ Optional for client, but infrequently used
  – Authenticates name of server
    ▪ But not if you are connecting to right place
More on TLS

• You need to be using it
  – If anything is sensitive
  – If you use passwords
• Many sites now turning on by default
  – Search ranking may also be influenced
Problems with TLS

• Many vulnerabilities
  – Often key choice
  – Often negotiation
• Usually one way authentication and no “authorization”.
• Man in the middle attacks
  – Superfish
  – Lenovo
• Data in open on client and server
Virtual Private Networks

• Problem
  – Lots of internet communication not protected by SSL/TLS or other means of encryption.
  – Network traffic in general is easily intercepted and read or modified.
  – Especially from Open Wifi, or hotel networks, but taps possible even from home or businesses.
  – Even SSL/TLS communications can be vulnerable depending on how connection initiated.
Virtual Private Networks

• Potential Solution
  – Create a tunnel from device to more central location in network
    ▪ Devices Business network
    ▪ Devices Home Network
    ▪ Third party network

• Issues
  – Traffic still unprotected once dumped onto primary network
  – But it protects traffic at the less secure endpoint.

• Other advantages
  – IP Address is not indicative of device location
Messaging Apps

- **WhatsApp**
  - Now owned by Facebook
  - Uses SSL encryption to communicate with server
  - Concerns – availability of data through the server?
  - Brazil Facebook head arrested for refusing to share WhatsApp data

- **Wickr**
  - A messaging app
  - Data is encrypted from device to device so not present on central servers.
SnapChat and ephemeral messaging

- SnapChat and similar ephemeral messaging apps imply that you can send a message that will disappear after it is viewed, or shortly thereafter.
- FTC recently took them to task for such claims.
- One is dependent on one's trust in the provider of the service.
- One can always videotape and take a photo of their own screen (using a second camera, etc).
Wikr is a better alternative

**HOW WICKR WORKS**

Our Peer-to-Peer Encryption does not rely on centralized private KDC for decryption

- ID and device undergo multiple rounds of salted cryptographic hashing using SHA256.
- Data at rest and in transit is encrypted with AES256.
- No password or password hashes leave the device.
- Messages and media are forensically wiped from the device after they expire.

---

**Message Sender**
- Encrypts locally
- All keys are randomly generated, unique for each message and user, and used only once then forensically destroyed

**Message Receiver**
- Decrypts locally
- Message bound to device

---

**PATENTED DIGITAL SECURITY BUBBLE**
- ECDH521
- Perfect Forward Secrecy
- Hardware Binding
- AES 256
- TLS
Anonymization

- Even if contents are protected, traffic analysis is still possible, providing information about what sites one visits.
- Tools are available that will hide your addresses
  - Proxies
  - Networks of Proxies – Onion Routing and TOR
Anonymizer and similar services

- Some are VPN based and hide IP addressed.
- Some of proxy based, where you configure your web browser.
- Need the proxy to hide cookies and header information provided by browser.
- You trust the provider to hide your details.
- Systems like TOR do better because you don’t depend on a single provider.
Onion Routing

• Layers of peer-to-peer anonymization.
  – You contact some node in the onion routing network
  – Your traffic is forward to other nodes in the network
  – Random delays and reordering is applied.
  – With fixed probability, it is forwarded on to its destination.

• TA requires linking packets through the full chain of participants.
  – And may be different for each association.
TOR

- Originally developed by US Navy to protect Internet communications
- The problem:
  - Internet packets have two parts – header and payload
  - Even if payload is encrypted, header is not
  - Header lists originator and destination nodes – all nodes along the way can read this information
- Law enforcement may not want it known they are visiting a site
Congratulations!
This browser is configured to use Tor.
You are now free to browse the Internet anonymously.

Test Tor Network Settings

Search securely with Startpage.

What Next?
Tor is NOT all you need to browse anonymously! You may need to change some of your browsing habits to ensure your identity stays safe.

Tips On Staying Anonymous »

You Can Help!
There are many ways you can help make the Tor Network faster and stronger:

• Run a Tor Relay Node »
• Volunteer Your Services »
• Make a Donation »

The Tor Project is a US 501(c)(3) non-profit dedicated to the research, development, and education of online anonymity and privacy. Learn more about The Tor Project »
TOR - Fundamentals

Source Node

Destination Node
Storage Encryption

- File Sharing (not necessarily encrypted)
- TrueCrypt
- PGP
File Encryption

• There are many tools and packages available to encrypt individual files or entire drives. Among these are the whole drive encryption discussed in the intro class, but software tools are also available.
• PGP file encrypt – part of the PGP package discussed earlier allows encryption of files or folders using the public key of an intended recipient (or yourself).
• TrueCrypt was for some time the best option for file encryption, but the last release removed the ability to encrypt files, and was accompanied by statements urging that it not be used. It is widely believed that the previous version is safe.