Privacy
Computer Security
Peter Reiher
December 11, 2014

Privacy

- Data privacy issues
- Network privacy issues
- Some privacy solutions

What Is Privacy?

- The ability to keep certain information secret
- Usually one's own information
- But also information that is "in your custody"
- Includes ongoing information about what you're doing

Privacy and Computers

- Much sensitive information currently kept on computers
  - Which are increasingly networked
- Often stored in large databases
  - Huge repositories of privacy time bombs
- We don't know where our information is

Privacy and Our Network Operations

- Lots of stuff goes on over the Internet
  - Banking and other commerce
  - Health care
  - Romance and sex
  - Family issues
  - Personal identity information
- We used to regard this stuff as private
  - Is it private any more?

Threat to Computer Privacy

- Cleartext transmission of data
- Poor security allows remote users to access our data
- Sites we visit save information on us
  - Multiple sites can combine information
- Governmental snooping
- Location privacy
- Insider threats in various places

Some Specific Privacy Problems

- Poorly secured databases that are remotely accessible
  - Or are stored on hackable computers
- Data mining by companies we interact with
- Eavesdropping on network communications by governments
- Insiders improperly accessing information
- Cell phone/mobile computer-based location tracking

Data Privacy Issues

- My data is stored somewhere
  - Can I control who can use it/see it?
- Can I even know who's got it?
- How do I protect a set of private data?
  - While still allowing some use?
- Will data mining divulge data "through the back door"?

Privacy of Personal Data

- Who owns data about you?
- What if it's really personal data?
  - Social security number, DoB, your DNA record?
- What if it's data someone gathered about you?
  - Your Google history or shopping records
  - Does it matter how they got it?

Protecting Data Sets

- If my company has (legitimately) a bunch of personal data,
- What can I/should I do to protect it?
  - Given that I probably also need to use it?
- If I fail, how do I know that?
  - And what remedies do I have?

Options for Protecting Data

- Careful system design
- Limited access to the database
  - Networked or otherwise
- Full logging and careful auditing
- Store only encrypted data
  - But what about when it must be used?
  - Key issues

Data Mining and Privacy

- Data mining allows users to extract models from databases
  - Based on aggregated information
- Often data mining allowed when direct extraction isn't
- Unless handled carefully, attackers can use mining to deduce record values

An Example of the Problem

- Netflix released a large database of user rankings of films
  - Anonymized, but each user had one random identity
- Clever researchers correlated the database with IMDB rankings
  - Which weren't anonymized
  - Allowed them to match IMDB names to Netflix random identities

Insider Threats and Privacy

- Often insiders need access to private data
  - Under some circumstances
- But they might abuse that access
- How can we determine when they misbehave?
- What can we do?

Local Examples

- Over 120 UCLA medical center employees improperly viewed celebrities' medical records
  - Between 2004-2006
- Two accidental postings of private UCLA medical data in 2011
- UCLA is far from the only offender

Encryption and Privacy

- Properly encrypted data can only be read by those who have the key
  - In most cases
  - And assuming proper cryptography is hazardous
- So why isn't keeping data encrypted the privacy solution?

Problems With Data Encryption for Privacy

- Who's got the key?
- How well have they protected the key?
- If I'm not storing my data, how sure am I that encryption was applied?
- How can the data be used when encrypted?
  - If I decrypt for use, what then?

A Recent Case

- Yahoo lost 450,000 user IDs and passwords in July 2012
  - The passwords weren't encrypted
  - Much less salted
- Password file clearly wasn't well protected, either
- Who else is storing your personal data unencrypted?

Steganography

- Another means of hiding data in plain sight
- In general terms, refers to embedding data into some other data
- In modern use, usually hiding data in an image
  - People have talked about using sound and other kinds of data

An Example

```
Transfer
$100 to my
savings
account
```

Run these through `outguess`

Voila!

The one on the right has the message hidden in it

How It Works

- Encode the message in the low order bits of the image
- Differences in these bits aren't human-visible
- More sophisticated methods also work
- Detected by looking for unlikely patterns
- Often foiled by altering images
- Steganography designers try to be robust against these problems

What's Steganography Good For?

- Used by some printer manufacturers to prove stuff came from them
- Stories of use by Al-Qaeda
  - No evidence of truth of stories
- Shady Rat attacks apparently used it to hide code to contact botnet servers
- Russian spies used it recently
- Most useful if opponents don't suspect you're using it

Steganography and Privacy

- If they don't know my personal data is in my family photos, maybe it's safe
- But are you sure they don't know?
  - Analysis of data used to store things steganographically may show that
- Essentially, kind of like crypto
  - But without the same level of mathematical understanding

Network Privacy

- Mostly issues of preserving privacy of data flowing through network
- Start with encryption
  - With good encryption, data values not readable
- So what's the problem?

Traffic Analysis Problems

- Sometimes desirable to hide that you're talking to someone else
- That can be deduced even if the data itself cannot
- How can you hide that?
  - In the Internet of today?

A Cautionary Example

- VoIP traffic is commonly encrypted
- Researchers recently showed that they could understand what was being said
  - Despite the encryption
  - Without breaking the encryption
  - Without obtaining the key

How Did They Do That?

- Lots of sophisticated data analysis based on understanding human speech
  - And how the application worked
- In essence, use size of encrypted packets and interarrival time
  - With enough analysis, got conversation about half right

Location Privacy

- Mobile devices often communicate while on the move
- Often providing information about their location
  - Perhaps detailed information
  - Maybe just hints
- This can be used to track our movements

Cellphones and Location

- Provider knows what cell tower you're using
- With some effort, can pinpoint you more accurately
- In US, law enforcement can get that information just by asking
  - Except in California

Other Electronic Communications and Location

- Easy to localize user based on hearing 802.11 wireless signals
- Many devices contain GPS nowadays
  - Often possible to get the GPS coordinates from that device
- Bugging a car with a GPS receiver not allowed without warrant
  - For now . . .

Implications of Location Privacy Problems

- Anyone with access to location data can know where we go
- Allowing government surveillance
- Or a private detective following your moves
- Or a maniac stalker figuring out where to ambush you . . .

Another Location Privacy Scenario

- Many parents like to know where their children are
- Used to be extremely difficult
- Give them a smart phone with the right app and it's trivial
- Good or bad?

A Bit of Irony

- To a large extent, Internet communications provide a lot of privacy
  - "On the Internet, no one knows you're a dog."
- But it's somewhat illusory
  - Unless you're a criminal

Why Isn't the Internet Private?

- All messages tagged with sender's IP address
- With sufficient legal authority, there are reliable mappings of IP to machine
  - ISP can do it without that authority
- Doesn't indicate who was using the machine
  - But owner is generally liable

Web Privacy

- Where we visit with our browsers reveals a lot about us
- Advertisers and other merchants really want that information
- Maybe we don't want to give it to them
  - Or to others
- But there are many technologies to allow tracking
  - Even to sites the tracker doesn't control

Do Not Track

- Wouldn't it be nice if we could ensure that web sites don't track us?
- Enter the Do Not Track standard
- A configurable option in your web browser
- Which, by enabling, you might think prevents you from being tracked

The Problems With Do Not Track

- First, it's voluntary
  - Web server is supposed to honor it
  - But will they?
- Second, and worse, it doesn't mean what you think it means
  - Based on current definitions of the option

What Do Not Track Really Means

- What it really means is “I’ll track you anyway”
- “But I won’t provide you anything helpful based on the tracking”
- So they know what you’re doing
  - And they do whatever they want with that data
- But you don’t see targeted ads
- So what’s the point of Do Not Track?
  - A good question

Some Privacy Solutions

- The Scott McNealy solution
  - "Get over it."
- Anonymizers
- Onion routing
- Privacy-preserving data mining
- Preserving location privacy
- Handling insider threats via optimistic security

Anonymizers

- Network sites that accept requests of various kinds from outsiders
- Then submit those requests
  - Under their own or fake identity
- Responses returned to the original requestor
- A NAT box is a poor man's anonymizer

The Problem With Anonymizers

- The entity running it knows who's who
- Either can use that information himself
- Or can be fooled/compelled/hacked to divulge it to others
- Generally not a reliable source of real anonymity

An Early Example

- A remailer service in Finland
- Concealed the actual email address of the sender
  - By receiving the mail and resending it under its own address
- Court order required owner of service to provide a real address
  - After which he shut down the service

Onion Routing

- Meant to handle issue of people knowing who you're talking to
- Basic idea is to conceal sources and destinations
- By sending lots of crypo-protected packets between lots of places
- Each packet goes through multiple hops

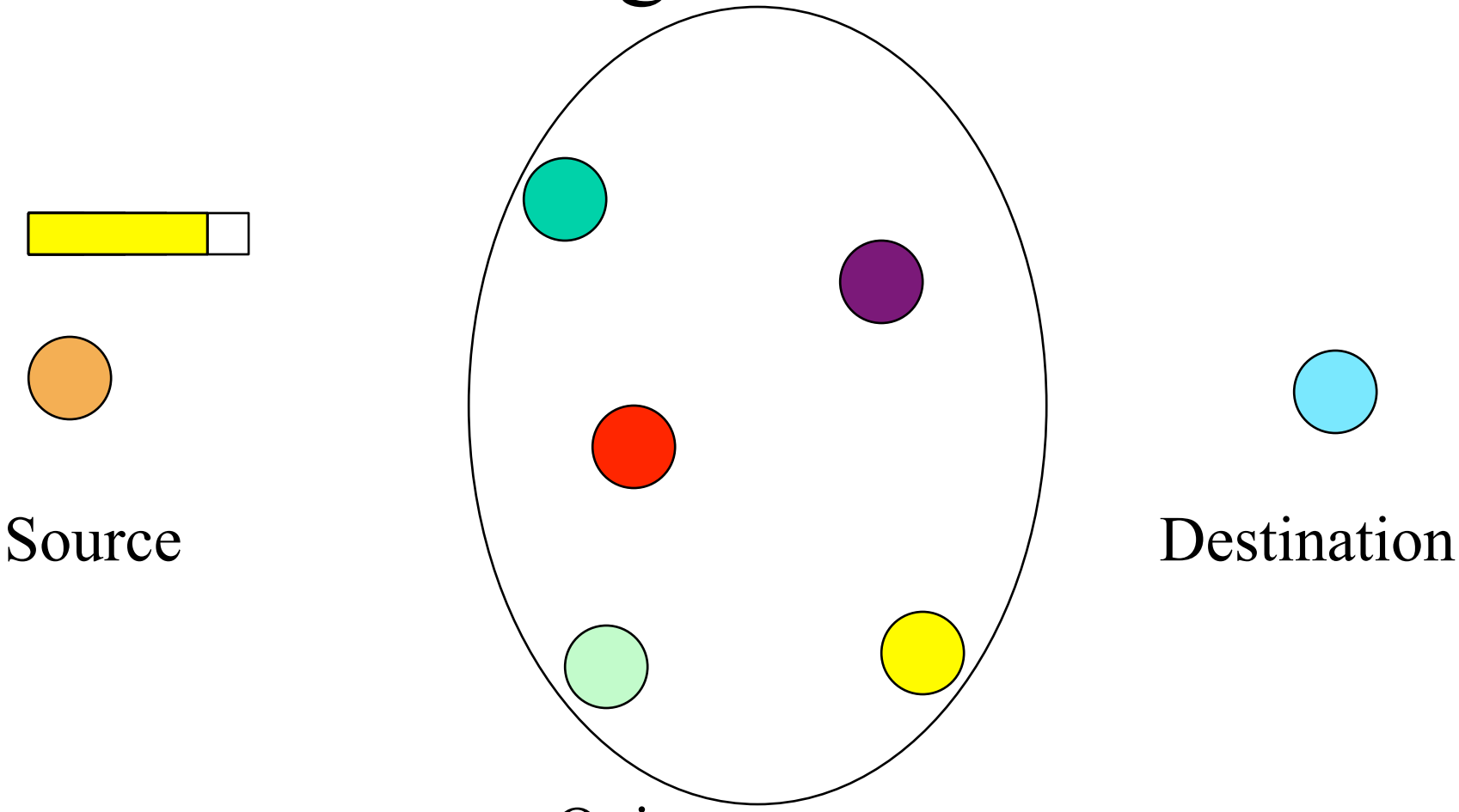
A Little More Detail

- A group of nodes agree to be onion routers
- Users obtain crypto keys for those nodes
- Plan is that many users send many packets through the onion routers
  - Concealing who's really talking

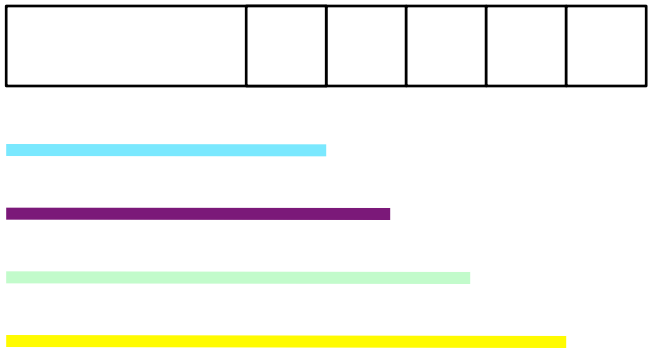
Sending an Onion-Routed Packet

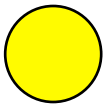
- Encrypt the packet using the destination's key
- Wrap that with another packet to another router
  - Encrypted with that router's key
- Iterate a bunch of times

In Diagram Form

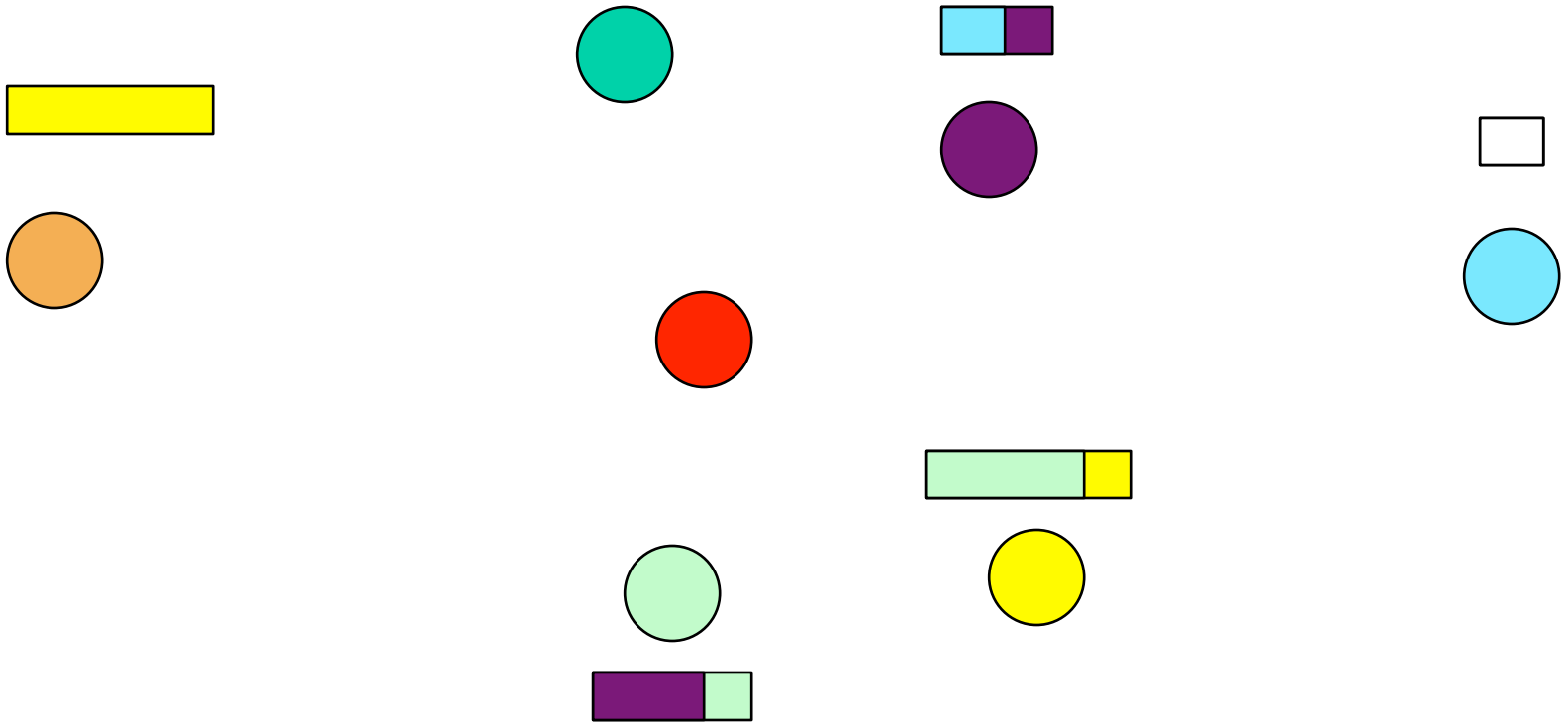


What's Really in the Packet



An unencrypted header to allow delivery to 

Delivering the Message



What's Been Achieved?

- Nobody improper read the message
- Nobody knows who sent the message
  - Except the receiver
- Nobody knows who received the message
  - Except the sender
- Assuming you got it all right

Issues for Onion Routing

- Proper use of keys
- Traffic analysis
- Overheads
  - Multiple hops
  - Multiple encryptions

Tor

- The most popular onion routing system
- Widely available on the Internet
- Using some of the original onion routing software
  - Significantly altered to handle various security problems
- Usable today, if you want to
- IETF is investigating standard for Tor

Why Hasn't Tor Solved This Privacy Problem?

- First, the limitations of onion routing
- Plus usability issues
  - Tor's as good as it gets, but isn't that easy to use
- Can't help if a national government disapproves
  - China and other nations have prohibited Tor's use

Can't I Surreptitiously Run Tor?

- Can't I get around government restrictions by just not telling them?
- No
  - Tor routers must know each others' identities
  - Traffic behavior of Tor routers "glows in the dark"
  - Tor developers keep trying

Privacy-Preserving Data Mining

- Allow users access to aggregate statistics
- But don't allow them to deduce individual statistics
- How to stop that?

Approaches to Privacy for Data Mining

- Perturbation
  - Add noise to sensitive value
- Blocking
  - Don't let aggregate query see sensitive value
- Sampling
  - Randomly sample only part of data

Preserving Location Privacy

- Can we prevent people from knowing where we are?
- Given that we carry mobile communications devices
- And that we might want location-specific services ourselves

Location-Tracking Services

- Services that get reports on our mobile device's position
  - Probably sent from that device
- Often useful
  - But sometimes we don't want them turned on
- So, turn them off then

But . . .

- What if we turn it off just before entering a “sensitive area”?
- And turn it back on right after we leave?
- Might someone deduce that we spent the time in that area?
- Very probably

Handling Location Inferencing

- Need to obscure that a user probably entered a particular area
- Can reduce update rate
  - Reducing certainty of travel
- Or bundle together areas
  - Increasing uncertainty of which was entered

So Can We Have Location Privacy?

- Not clear
- An intellectual race between those seeking to obscure things
- And those seeking to analyze them
- Other privacy technologies (like Tor) have the same characteristic

The NSA and Privacy

- 2013 revelations about NSA spying programs changed conversation on privacy
- The NSA is more heavily involved in surveillance than previously believed
- What are they doing and what does that mean for privacy?

Conclusion

- Privacy is a difficult problem in computer systems
- Good tools are lacking
  - Or are expensive/cumbersome
- Hard to get cooperation of others
- Probably an area where legal assistance is required