

---

# Voluntary Disclosures



---

## Give-Aways

- Not all privacy problems are due to evil social networking sites and web sites
- Some information is given away
- People don't always realize the consequences of what they say online

---

## War 2.0

“Israel’s Army Radio reported on Wednesday that a raid on suspected militants in the West Bank planned for Wednesday was called off by the country’s military because a soldier posted details of the operation on Facebook.

“The Israeli newspaper Haaretz explained that the soldier posted a status update letting friends know that his unit was preparing to go to a West Bank village near Ramallah: On Wednesday we clean up Qatanah, and on Thursday, god willing, we come home, the soldier wrote.”

---

## Inappropriate Publications Aren't New

- During World War II, the *Chicago Tribune* published a story suggesting that the U.S. was reading Japan's naval codes — which was true. . .
- (The Japanese never saw the story)
- What's different now?

---

## The Difference

- No gatekeeper
- No single person to hold accountable (Roosevelt wanted to use the Marines to shut down the *Chicago Tribune* and/or wanted to charge the publisher with treason)
- It's the same thing that makes Web 2.0 useful for other reasons — publishing is now decentralized

---

# Usenet

- The decentralized publication problem started with Usenet
- People engaged in discussions would often post proprietary material
- Most companies required preclearance for publication — but was Usenet “publishing” or “talking with friends”?
- The social dynamic was the latter, but what was posted was rapidly distributed worldwide.

---

## The Web Made it Worse

- Inappropriate content was now permanent
- Search engines could find it
- Better yet — search engines could find corporate markings for proprietary documents
- Home in on the good stuff...

---

## Unofficial Collaboration Channels

- Often, official mechanisms for sharing data don't work well
- Simple solution: put the files on a web site and *assume* that no one else will see them
- But web servers will auto-list directories and search engines can find things...

---

## Autolisting Directories

### Index of /~smb/foo

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
 <a href="#">Parent Directory</a>		-	
 <a href="#">test1.txt</a>	08-Mar-2010 22:44	7	
 <a href="#">test2.txt</a>	08-Mar-2010 22:44	7	

---

*Apache Server at [www.cs.columbia.edu](http://www.cs.columbia.edu) Port 80*

---

## Peer-to-Peer Filesharing

- Peer-to-peer file-sharing programs are designed to share files with minimal effort
- But what files?
- The goal — let's be honest — was to share MP3s and movies
- Some programs' default configuration shared all file types in certain directories.
- Do people put their music in **My Documents**?

---

## Apparently, Some Do...

- Empirically, we do see documents on file-sharing networks
- In February, the FTC notified about 100 organizations that sensitive customer and employee data was being shared that way
- “They might accidentally choose to share drives or folders that contain sensitive information, or they could save a private file to a shared drive or folder by mistake, making that private file available to others. In addition, viruses and other malware can change the drives or folders designated for sharing, putting private files at risk. As a result, instead of just sharing music, a user’s personal tax returns, private medical records or work documents could end up in general circulation on P2P networks.” (FTC)

---

## A New Law?

- There's a bill before Congress to require warnings about P2P programs
- Program must provide “clear and conspicuous notice that such program allows files on the protected computer to be made available for searching by and copying to one or more other computers”
- Such programs must not make themselves unduly hard to remove

---

## You Can't Call it Back

- Once something is on the Net, it's probably there forever
- Email with the text "**XXX wishes to recall the message  
"YYYY"**" is ineffective on anything other than Microsoft Exchange
- In fact, it only draws more attention to the message...

---

## Sexting

- People – frequently teenage girls — take “suggestive” photos of themselves with their phones and send them to their (boy)friends
- The recipients display lots of class and circulate the pictures to several dozen of their closest, most trusted friends
- These “friends”, of course, share them further
- One result in several jurisdictions: threatened or actual child pornography charges

---

## It's Not Just Teens

- In 2001, a British attorney at a large multinational law firm forwarded a complimentary private email from his girlfriend — he wanted to brag about his “prowess”
- One of his friends sent it further, saying that he felt “honour bound” to redistribute it
- The email rapidly became very public
- The attorney was fired; several of the forwarders were disciplined

---

## What Happened?

- Imagine a wandering bard, spreading a “naughty” rumor wherever he goes
- Imagine a monk in a scriptorium making several copies of such stories and sending them out by courier
- Someone takes a letter to the copier room, makes lots of copies, and mails them
- What is the bandwidth in recipients/second?
- A quantitative difference in speed has become a qualitative difference

---

## Can We Create Time-Limited Email?

- General approach: encrypt the email with some key; destroy the key after a certain interval
- To read the email, the recipient first has to either retrieve the key or upload the mail to a decryption server
- In either case, the message is displayed on the recipient's screen, which can be copied

---

## The Converse Problem

- On the other hand, creating a message which can't be read until a certain time is (in principle) easy
- Launch satellites with private cryptographic keys into the L4 and L5 Earth-Sun Lagrange points
- Multiply-encrypt the message, alternating between the public keys for each of the two satellites
- Each leg — they're 8 light-minutes from earth, and 13.8 light-minutes apart — can do one decryption; there's no way to talk faster than the speed of light

---

# Social Networks

- Facebook et al. are personal places
- People post things about themselves and their friends
- Such postings can follow them around

---

## What's the Problem?

- Secondary use — personal data is being used when judging employment suitability
- Persistence of data — data has a long lifetime
- The “Wayback Machine”: [www.archive.org](http://www.archive.org)

# The Wayback Machine

Jan 01, 1996 - Sep 10, 2009				
2003	2004	2005	2006	2007
0 pages	0 pages	17 pages	20 pages	6 pages
		<a href="#">Jan 11, 2005</a> * <a href="#">Jan 11, 2005</a> * <a href="#">Feb 12, 2005</a> * <a href="#">Feb 12, 2005</a> * <a href="#">Mar 09, 2005</a> <a href="#">Sep 07, 2005</a> * <a href="#">Sep 10, 2005</a> <a href="#">Sep 14, 2005</a> <a href="#">Sep 17, 2005</a> <a href="#">Sep 23, 2005</a> <a href="#">Oct 18, 2005</a> <a href="#">Oct 28, 2005</a> <a href="#">Nov 09, 2005</a> <a href="#">Dec 03, 2005</a> * <a href="#">Dec 10, 2005</a> <a href="#">Dec 12, 2005</a> <a href="#">Dec 31, 2005</a>	<a href="#">Jan 01, 2006</a> <a href="#">Jan 02, 2006</a> <a href="#">Jan 09, 2006</a> <a href="#">Feb 09, 2006</a> <a href="#">Apr 25, 2006</a> <a href="#">Jun 19, 2006</a> <a href="#">Aug 20, 2006</a> <a href="#">Aug 30, 2006</a> <a href="#">Aug 31, 2006</a> <a href="#">Sep 01, 2006</a> <a href="#">Sep 07, 2006</a> * <a href="#">Sep 07, 2006</a> * <a href="#">Sep 11, 2006</a> <a href="#">Sep 20, 2006</a> <a href="#">Sep 21, 2006</a> * <a href="#">Sep 21, 2006</a> * <a href="#">Sep 23, 2006</a> <a href="#">Sep 24, 2006</a> * <a href="#">Oct 17, 2006</a> * <a href="#">Dec 10, 2006</a>	<a href="#">Feb 05, 2007</a> <a href="#">May 03, 2007</a> <a href="#">May 05, 2007</a> <a href="#">May 08, 2007</a> <a href="#">Jun 09, 2007</a> * <a href="#">Aug 19, 2007</a> *

---

## More Subtle Leaks

- Association patterns leak
- “Birds of a feather. . . .”
- Democrats (mostly) friend Democrats, Republicans (mostly) friend Republicans, gays mostly friend gays, etc.

---

## Protecting Facebook Data

- Facebook has elaborate privacy controls
- Many people are unaware of them
- They're not easy to use well
- Facebook says that 20M users changed some settings — but they have 350M users (NY Times)
- One nice feature: it's possible to see how anyone on your friend list will see your profile

---

## They Keep Changing Things

- Facebook frequently changes its policies and mechanisms
- New defaults are more open
- This is a conscious decision by Facebook, which thinks that people want more openness
- Also: how will Facebook make money?

---

# EXIF Tags

- JPGs contain a lot of metadata

- Some is about the photo:

Exposure Time: 1/60 sec

F-Number: f/4.5

Exposure Program: Not Defined

Exposure Bias: 0 EV

Metering Mode: Pattern

Light Source: Unknown

Flash: Flash, Auto, Return Detected

Focal Length: 44.00 mm

- But it can also include the camera model and serial number and GPS location

---

## Conclusions

- People often give away information without knowing it
- There's no one cause: it can be hidden (EXIF data) or newly-learnable (link analysis), or it can be carelessness
- Social networking sites default to openness
- Is there a problem?