Course Program

Monday, June 27
• 09:30-10:30 Introduction, Claudia Diaz
• 11:00-12:30 Privacy Enhancing Cryptographic Protocols, Markulf Kohlweiss (Invited Talk)
• 14:00-15 Recent insights in HCI and user research on location based privacy, David Geerts
• 16:00-17:00 Technical Aspects of Location Privacy, Claudia Diaz

Tuesday, June 28
• 09:00-10:30 Web data mining and privacy: foes or friends? Bettina Berendt
• 11:00-12:30 The surge of data analytics and the right to know how one is being profiled, Mireille Hildebrandt (Invited Talk)
• 14:00-15:30 Location privacy in the European data protection legal framework: when your smartphone is getting "too" smart, Eleni Kosta
• 16:00-17:30 The challenge of privacy and requirements engineering in information systems, Seda Gürses
• 17:30 Discussion speakers and participants
Introduction

Claudia Diaz
K.U.Leuven ESAT/COSIC

June 27, 2011
“offline world” vs “online world”

- face-to-face conversation
- letters in the post
- Knowing who your friends are
- Paying with cash
- Following your movements
- Learning your shopping profile
- Papers in physical archives

- Easy/cheap to collect, store search, and process
- Easy to copy/disseminate, but hard to destroy
- Easy to aggregate, make profiles and inferences
- Information never forgotten
Information taken out of context

The Officer Who Posted Too Much on MySpace

By AMY WIG 
Published March 10, 2009

In pictures, Vaughan Etienne is a champion bodybuilder of surreal musculature. In conversation, he is polite and thoughtful.

And in the looking glass of his computer screen, he becomes a man of fierce, profane views on how to keep law and order. A few weeks ago, he posted a description of his mood on a MySpace account. “Devious,” he wrote.

The next day, a man accused of carrying a loaded gun would go on trial in State Supreme Court in Brooklyn — and in large part, the case rested on the credibility of Vaughan Etienne, bodybuilder, Internet user and arresting officer.

What seemed like a simple gun possession case became an undeclared war over reality. Was Officer Etienne a diligent cop who found a gun after chasing an ex-convict weaving through traffic on a stolen motorcycle? Or was his story a “devious” facade in keeping with the ruthless character he revealed on social network Web sites?

“You have your Internet persona, and you have what you actually do on the street,” Officer Etienne said on Tuesday. “What you say on the Internet is all bravado talk, like what you say in a locker room.”

Except that trash talk in locker rooms almost never winds up preserved on a digital server somewhere, available for autoplay. The
Identity theft

Personal data privacy 'at risk'

Millions of people are leaving themselves open to identity theft when using social networking websites, according to the consumer group Which?

Members of sites such as Facebook can join large networks which reveal personal information to thousands of others on the network.

Which? says people are at a greater risk of being targeted by fraudsters than they think.

On average, UK residents' details are held on about 700 databases.

Which? says that fraudsters can use the internet to gather personal information which could then be used to trick people into revealing PIN numbers and other security information.

These could then be used by conmen to apply for credit cards or loans in somebody else's name.

Burglars could also benefit from such information, it says.
Spear phishing

**Spear-Phishing:**
Social Networks open to harvesting

In a recent article, I outlined the problems with increased breaches of security on sites as huge as MySpace and Facebook, not to mention sites that are supposed to be geared specifically towards businesses. While there is still a digital DMZ around most enterprise, those who would compromise your users and security are going after these so-called soft targets with a new twist on the old social engineering methods to extract information and gain access to your resources. Here is what you should know about this latest ploy!

**Social Engineering**
Social engineering is nothing new. It has been around since before there were PCs. An individual talks their way into an office, claiming to be there to solve a problem from someone, working on the phones, picking up a file, etc. Social engineering is alive and well and remains the most effective hacking technique according to Aaron Higbee, a managing partner and co-founder of the Intrepidus Group. Its use assists in proliferation of malware, spam, mass phishing, botnets and even more advanced hacking techniques.

Today’s social engineering threats have progressed beyond the scope of infamous hacker Kevin Mitnick, who earned his reputation and subsequent jail time by using his ability to find soft targets within an organization and schmoozing them into providing information that would allow him access to their systems and network. Now out, Kevin Mitnick runs Mitnick Security Consulting LLC, a computer security consultancy firm.

**Spear-Phishing?**
Ok, so you have heard of Phishing, the electronic equivalent of throwing a hook and bait into a lake. Instead of a fish, the target of Phishing is information. So what is spear-phishing?
Stalking

Social Networking Sites: A Bonanza for Stalkers?

"Vengeance will be mine....," declared a defiant message on MySpace.com. "I should have killed you all when I had a gun and some drugs." This violent monologue, one of several postings on the writer's site, threatened his ex-wife, who had fled the state to escape his abuse. In postings on other sites, he demanded photos of his family and warned that if he didn't get to see the kids, "it isn't going to be real good, because I'm gonna see them whether you let me or not."[1]

The increasing use of MySpace to threaten and stalk victims raises many important questions. Do social networking sites enable stalking? What recourse do victims have when these sites are used to stalk? And what tools can help block the use of these sites to stalk?

What Are Social Networking Sites?

Social networking sites such as MySpace and Facebook are virtual communities where people with mutual interests meet on-line to share information and build relationships. Site visitors can chat, debate, network, and socialize. On many sites, members may post details about themselves—photos; educational backgrounds; favorite books, movies, and music; and relationship status. Others sites promote business, activism, networking, counseling, socializing, or many types of recreational interests. Sites such as MySpace, Facebook, Friendster, and Xanga have attracted millions of members, particularly among teenagers and young adults.
Predators on Social Networks

Sexual Predators and Child Molesters Find Easy Prey

From Tony Bradley, CISSP ISSAP, former About.com Guide

Social networking is all the rage. Various websites have sprung up for the sole purpose of providing a place for users to express themselves, share with like-minded individuals, discover new things, and communicate with others. Even I have a Myspace profile and a LinkedIn profile.

The concept is so popular that even the 400-pound gorillas of the Web have jumped on the bandwagon. MySpace was snatched up by Rupert Murdoch’s News Corp. Google has Orkut. Yahoo tried Yahoo 360, and is now beta-testing their new social network dubbed Mash. Microsoft just bought into a large stake of Facebook.

The concept of social networking has also been extended to other areas. For example, Youtube (also picked up by Google), provides users with the ability to express their creativity, network, rate their favorite video clips, etc. Some sites like Flickr, DropShots, or PhotoBucket provide users with the ability to post and share photos and family videos.
When you browse to a website

“Privacy Diffusion on the Web: A Longitudinal Perspective” B. Krishnamurthy and C. Wills
Study of 1200 popular Web sites
NHS.uk allowing Google, Facebook, and others to track you

The NHS is allowing Google, Facebook, and others to track your http://www.nhs.uk/browsing habits, regardless of the fact that people use the page to seek medical advice. It was recently pointed out to me that the NHS Choices website’s social features include the Facebook Like button (see e.g. the page on Testicular Cancer). Due to the fact that the standard method of Facebook Like button deployment is intrusive to say the least, I thought I would look into identifying which third party companies have been given permission to track users on NHS Choices, and my results are rather disconcerting.

In short there are four third-party, advertising/tracking companies which are informed every time a user visits one of the “conditions pages” on the NHS Choices website. These listed below, all get to make a call from the user’s browser, in turn allowing the four companies to access their cookies, tracking the users (explained in a previous blog post of mine, and in Bala’s research). This means, that if one has ever logged into a Google account, or a Facebook account and then visits one of the pages on the NHS site, the company will then know that their user X was just looking at a page about condition Y on the NHS website.
Hidden inside Ashley Hayes-Beatty's computer, a tiny file helps gather personal details about her, all to be put up for sale for a tenth of a penny.

The file consists of a single code—4c812db29272995e5416a323e79bd37—that secretly identifies her as a 28-year-old female in Nashville, Tenn.

When a user like you logs onto the internet ...

The code knows that her favorite movies include "The Princess Bride," "50 First Dates" and "10 Things I Hate About You." It knows she enjoys the "Sex and the City" series. It knows she browses entertainment news and likes to take quizzes.

"Well, I like to think I have some mystery left to me, but apparently not!" Ms. Hayes-Beatty said when told what that snippet of code reveals about her. "The profile is eerily correct."

Ms. Hayes-Beatty is being monitored by Lotame Solutions Inc., a New York company that uses sophisticated software called a "beacon" to capture what people are typing on a website—their comments on movies, say, or their interest in expecting and pregnancy. Lotame packages that data into profiles about individuals, without determining a person's name, and sells the profiles to companies seeking customers. Ms. Hayes-Beatty's tastes can be sold wholesale (a batch of movie lovers is $1 per thousand) or customized (26-year-old Southern fans of "50 First Dates").
By EMILY STEEL and JULIA ANGWIN

(Please see Corrections & Amplifications item below)

You may not know a company called [x+1] Inc., but it may well know a lot about you.

**One Smart Cookie**

New York ad company [x+1] made predictions about users based on just one click on a website. Read more about the users, see the code transmitted and review the company’s assumptions.

From a single click on a web site, [x+1] correctly identified Carrie Isaac as a young Colorado Springs parent who lives on about $50,000 a year, shops at Wal-Mart and rents kids’ videos. The company deduced that Paul Bouliffard, a Nashville architect, is childless, likes to travel and buys used cars. And [x+1] determined that Thomas Ermey, a Colorado building contractor, is a skier with a college degree and looks like he has good credit.

The company didn’t get every detail correct. But its ability to make snap assessments of individuals is accurate enough that Capital One Financial Corp. uses [x+1]’s calculations to instantly decide which credit cards to show first-time visitors to its website.

In short: Websites are gaining the ability to decide whether or not you’d be a good customer, before you tell them a single thing about yourself.
Life insurers are testing an intensely personal new use for the vast dossiers of data being amassed about Americans: predicting people's longevity.

Insurers have long used blood and urine tests to assess people's health—a costly process. Today, however, data-gathering companies have such extensive files on most U.S. consumers—online shopping details, catalog purchases, magazine subscriptions, leisure activities and information from social-networking sites—that some insurers are exploring whether data can reveal nearly as much about a person as a lab analysis of their bodily fluids.
A research project of the Electronic Frontier Foundation

Panopticlick
How Unique — and Trackable — Is Your Browser?

Is your browser configuration rare or unique? If so, web sites may be able to track you, even if you limit or disable cookies.

Panopticlick tests your browser to see how unique it is based on the information it will share with sites it visits. Click below and you will be given a uniqueness score, letting you see how easily identifiable you might be as you surf the web.

Only anonymous data will be collected by this site.

TEST ME

A paper reporting the statistical results of this experiment is now available: How Unique Is Your Browser?, Proceedings of the Privacy Enhancing Technologies Symposium (PETS 2010), Springer Lecture Notes in Computer Science.
Your Apps Are Watching You

A WSJ Investigation finds that iPhone and Android apps are breaching the privacy of smartphone users

By SCOTT THURM and YUKARI IWATANI KANE

What we found on one app

The iPhone version of music app Pandora sent information to eight trackers. It sent location data to seven of these, a unique phone ID to three and demographic data to two.

Few devices know more personal details about people than the smartphones in their pockets: phone numbers, current location, often the owner's real name—even a unique ID number that can never be changed or turned off.

These phones don't keep secrets. They are sharing this personal data widely and regularly, a Wall Street Journal investigation has found.

An examination of 101 popular smartphone "apps"—games and other software applications for iPhone and Android phones—showed that 56 transmitted the phone's unique device ID to other companies without users' awareness or consent. Forty-seven apps transmitted the phone's location in some way. Five sent age, gender and other personal details to outsiders.
Gaydar Algorithm Outs Facebook Users

By Susannah F. Locke  Posted 09.21.2009 at 12:27 pm  9 Comments

What are your friends saying about you? Online social networks like this Facebook one might reveal more about you than you think jurvetson (CC licensed)

A pair of MIT students claim that they have created an algorithm that outs gay members of Facebook by analyzing the sexual orientations of their networks of friends.
Not just the web...

- Pay-as-you drive applications
- Smart energy meters
- Airport scanners
- Loyalty cards
- “Smart” environments
- Genome studies
- ...

June 27, 2011
Defining Privacy
What is privacy?

- Abstract and subjective concept, hard to define
  - Dependent on cultural issues, study discipline, stakeholder, context

- Popular definitions:
  - “The right to be let alone”
    - Focus on freedom from intrusion
  - “Informational self-determination”
    - Focus on control
  - “The freedom from unreasonable constraints on the construction of one's own identity”
    - Focus on autonomy
Solove’s taxonomy of privacy

- **Information Collection**
  - Surveillance
  - Interrogation

- **Information Processing**
  - Aggregation
  - Identification
  - Insecurity
  - Secondary Use
  - Exclusion

- **Information Dissemination**
  - Breach of Confidentiality
  - Disclosure
  - Exposure
  - Increased Accessibility
  - Blackmail
  - Appropriation
  - Distortion

- **Invasion**
  - Intrusion
  - Decisional Interference

June 27, 2011
Data protection (EU)

- Data collected for specific and legitimate **purpose**
- **Proportional**: adequate, relevant and not excessive (data minimization)
- With the subject’s awareness and **consent**
  - Unless data is necessary for...
- Data subject’s right to access, correct, delete her data
- Data security: integrity, confidentiality of the data
Technical privacy properties

- Anonymity
- Pseudonymity
- Unlinkability
- Unobservability
- Plausible deniability
- Location privacy
- Censorship resistance
- ...

June 27, 2011
Useful distinctions?

- What is the concern?
  - Privacy towards other individuals relevant to our lives
  - Concentration of power in organizations who perform mass data collection and analysis

- What type of data?
  - Gives: the content you upload or share
  - Give-offs: clicks, search queries, behavior
Privacy models:
soft and hard privacy
Trust-based privacy

- **System model**
  - Data subject provides her data
  - Data controller responsible (trusted) for its protection
  - One or several data processors

- **Threat model**
  - External parties, errors, malicious insider
Trust-based privacy

- Controller/processors: main “users” of security technologies
- Policies, access control, audits (liability)
Trust-based privacy

- Data subject has already lost control of her data
- In practice, very difficult for data subject to verify how her data is collected and processed
- DP does not apply to “anonymized” data
- secrecy of algorithms
Trust-based privacy

- Data subject has already lost control of her data
  - In practice, very difficult for data subject to verify how her data is collected and processed
  - DP does not apply to “anonymized” data
  - secrecy of algorithms
  - Need to trust data controllers (honesty, competence) and hope for the best
restitution and audit, the total cost per compromised record has risen from $182 in 2006, to $204 in 2009. As society moves more and more digital, the threat of a breach becomes more evident. In 2010, there were 662 data breaches reported constituting over $16 Billion dollars in costs. Many of these costs came from lost business due to customer
Problems of trust-based privacy

- Data minimization (proportionality) often ignored
- Informed consent?

- Trust assumptions may not be realistic
  - Incompetence
  - Malicious insiders
  - Purpose (function creep)
  - Cost of securing the data
  - Incentives?

- Technologically enforced?
  - A “trusted” entity can undetectably violate the security policies
  - Weak DP enforcement, low penalties

- How can you check that your data is not being abused?
Facebook CEO Mark Zuckerberg and his company are suddenly facing a big new round of scrutiny and criticism about their cavalier attitude toward user privacy.

An early instant messenger exchange Mark had with a college friend won't help put these concerns to rest.

According to SAI sources, the following exchange is between a 19-year-old Mark Zuckerberg and a friend shortly after Mark launched The Facebook in his dorm room:

**Zuck**: Yeah so if you ever need info about anyone at Harvard

**Zuck**: Just ask.

**Zuck**: I have over 4,000 emails, pictures, addresses, SNS

[Redacted Friend's Name]: What? How'd you manage that one?

**Zuck**: People just submitted it.

**Zuck**: I don't know why.

**Zuck**: They "trust me"

**Zuck**: Dumb fucks.

Brutal.
Hard Privacy (PETs)

- System model
  - Subject provides as little data as possible
  - Reduce as much as possible the need to “trust” other entities

- Threat model
  - Strategic adversary with certain resources motivated to breach privacy (similar to security systems)
  - Adversarial environment: communication provider, data holder
Hard Privacy (PETs)

- Subject is an active security "user"
- Goal (data protection): data minimization
Two main approaches

- Anonymity
  - Service provider can observe access to the service
  - Cannot observe the identity of the user

- Oblivious Transfer (OT) / Private Information Retrieval (PIR)
  - Service provider can identify user
  - Cannot observe details of the access to the service
    - Which records were accessed
    - Which search keywords were used
    - Which content was downloaded
    - ...

- All parties have assurance that the other participants in the protocol are cannot cheat
Conclusions

- Increasing number of applications with implications for privacy: which will be the long-term effects?
- Variety of threats: privacy is not just about your mother/employer seeing your drunken pictures!
- Two different paradigms for developing privacy-friendly applications
  - They can/must be combined
- Trust-based privacy is the state of the art
  - Hidden costs of securing the data silos
- PETs can reconcile aggressive data minimization and service integrity guarantees
  - Active research but poor deployment