Privacy in the Information Society

Claudia Diaz

Katholieke Universiteit Leuven, Belgium
claudia.diaz@esat.kuleuven.be
http://homes.esat.kuleuven.be/~cdiaz
November 10, 2009
Outline

- Context: information processing and uniqueness
- Example applications with implications for privacy
- The privacy debate
- Privacy technologies
- Conclusions
Information processing

- manual processing ($10^2$)
- mechanical processing ($10^4$)
- mainframe ($10^5$)
- PCs and LANs ($10^7$)
- Internet and mobile ($10^9$)
- the Internet of things, ubiquitous computing, pervasive computing, ambient intelligence ($10^{12}$)
Information storage and transmission

- 2010: 1 Zettabyte/year added to the digital universe; this corresponds to 400 million hard drives with a capacity of 2.5 Terabyte
- 2010: US internet traffic will grow to 1 Zettabyte/month (today: 4 Exabyte)

Exabyte

<table>
<thead>
<tr>
<th>Year</th>
<th>Information</th>
<th>Available Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>2006</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>2007</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>2008</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>2009</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>2010</td>
<td>1200</td>
<td>1200</td>
</tr>
</tbody>
</table>

Megabyte $10^6$
Gigabyte $10^9$
Terabyte $10^{12}$
Petabyte $10^{15}$
Exabyte $10^{18}$
Zettabyte $10^{21}$

- photo: 1 Mbyte
- song: 50 Mbyte
- movie: 4.7 Gbyte
- 95 yrs: $3.10^9$ seconds
- life movie: 3 Petabyte
Uniqueness

- Physics and electronics (accidental)
  - Radio fingerprinting: unique pattern of each wireless
- Physics and electronics (deliberate)
  - MAC address
  - Pentium III Processor Serial Number
  - Certificate in e-ID card
- Human: biometry
  - Fingerprint, iris, face, DNA
- Human: behavior
  - Typing, gait, writing style
Intelligent processing
uniqueness + connectivity + processing power

- Create “big brother” for specific purposes
  - protecting children
  - road pricing and congestion control
  - public transport
  - car pool
  - prosecution of crime
  - health emergencies (home monitoring)
  - ambient intelligence
  - copyright infringements

- Much of the data used in these applications relates to people

- Individual applications are legitimate

- Cost effective
  - use of centralized backend systems
  - more information leads to more economic efficiency

- Long term incentive for integrating solutions and function creep

inexpensive mass surveillance
Exclusive: U.S. Spies Buy Stake in Firm That Monitors Blogs, Tweets

By Noel Shankman | October 19, 2009 | 12:03 pm | Categories: Info War, Spies, Secrecy and Surveillance

America's spy agencies want to read your blog posts, keep track of your Twitter updates — even check out your book reviews on Amazon.

In-Q-Tel, the investment arm of the CIA and the wider intelligence community, is putting cash into Visible Technologies, a software firm that specializes in monitoring social media. It's part of a larger movement within the spy services to get better at using "open source intelligence" — information that's publicly available, but often hidden in the flood of TV shows, newspaper articles, blog posts, online videos and radio reports generated every day.

Visible crawls over half a million web 2.0 sites a day, scraping more than a million posts and conversations taking place on blogs, online forums, Flickr, YouTube, Twitter and Amazon. (It doesn't touch closed social networks, like Facebook, at the moment.) Customers get customized, real-time feeds of what's being said on these sites, based on a series of keywords.

"That's kind of the basic step — get in and monitor," says company senior vice president Blake Cahill.
From The Times

July 6, 2009

Wife of Sir John Sawers, the future head of MI6, in Facebook security alert

Michael Evans, Defence Editor

Diplomats and civil servants are to be warned about the danger of putting details of their family and career on social networking websites. The advice comes after the wife of Sir John Sawers, the next head of MI6, put family details on Facebook — which is accessible to millions of internet users.

Lady Sawers disclosed details such as the location of the London flat used by the couple and the whereabouts of their three children and of Sir John’s parents. She put no privacy protection on her account, allowing any of Facebook’s 200 million users in the open-access London network to see the entries.

Lady Sawers’ half-brother, Hugo Haig-Thomas, a former diplomat, was among those
Some of the threats of surveillance

- Chilling effect on free speech
  - Expressing dissent
  - Organized movements

- Increase asymmetry of power
  - Those with information resources even more at an advantage (checks and balances?)
  - Manipulation, blackmail

- Loss of control of personal information (and image)

- Profiling
  - Discrimination

- Information stays forever (what you said at 15 might be looked at when you are 30)
  - Second opportunities?
Some example applications with implications for privacy
How many ways have you been located today?

- cell phone (turned on?)
- laptop computer
- credit card at the gas station
- bank card in the ATM machine
- driving through a monitored intersection
- security camera at the supermarket
- scan badge to enter a building
- pass a Bluetooth-enabled printer
“Chattering” devices

- RFID
- Bluetooth/Zigbee
- WLAN
- WiMAX
- 2G/GSM
- 3GSM
- GPS/Glonass/Galileo
Location Based Services

- Location-based traffic monitoring and emergency services
  - e-Call, traffic congestion control
  - Pay-as-you-drive
- Location finder:
  - Where is the nearest restaurant, gas station,...
- Social applications
  - Geotagged Twitter
  - Google Latitude
- Ubiquitous environments

Gartner Revenue of LBS:
- 2008: 998.3 M$
- 2009: 2.2 B$
Why is this a problem?

- Do you want to be seen at certain locations?
  - Abortion clinic, AIDS clinic, business competitor, or political headquarters (Google Street View)
- What can be automatically inferred about a person based on location?
  - Child: at school location during school hours
  - Home and work address
  - Movement routines, changes in routine
- Personal safety (e.g., stalking, child abduction)
- Spionage (high government officials, journalists, CEOs, military…)}
Darling admits Revenue loss of 25 million personal records
Lost: Two discs, 25 million accounts

By John Oates • Get more from this author

Posted in Government, 20th November 2007 16:22 GMT
Free whitepaper – The human factor in laptop encryption

UK Identity Crisis Alistair Darling told the House of Commons this afternoon that a police investigation has been launched into how Her Majesty's Revenue and Customs has lost child benefit records relating to 25 million people.

Records for 25 million people, relating to child benefit payments for 7.25 million families, were sent using the HMRC's own postal system, called grid, but never arrived.
Clarkson stung after bank prank

TV presenter Jeremy Clarkson has lost money after publishing his bank details in his newspaper column.

The Top Gear host revealed his account numbers after rubbing the furor over the loss of 25 million people's personal details on two computer discs.

He wanted to prove the story was a fuss about nothing.

But Clarkson admitted he was "wrong" after he discovered a reader had used the details to create a £500 direct debit to the charity Diabetes UK.

Clarkson published details of his Barclays account in the Sun newspaper, including his account number and sort code. He even told people how to find out his address.

"All you'll be able to do with them is put money into my account. Not take it out. Honestly, I've never known such a palaver about nothing," he told readers.
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.
The Officer Who Posted Too Much on MySpace

By JIM OWYER
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 subpoena.
Gaydar Algorithm Outs Facebook Users
By Susannah P. Locke  Posted 09/21/2009 at 12:27 pm  7 Comments

What are your friends saying about you? On the social networks like this Facebook one might reveal more about you than you think. [furnetsam](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.
Why is this a problem?

- Information taken out of context
  - Social problems (with employer, spouse, friends, etc.)
- Identity fraud
  - Getting a credit card on your name
- Spear phishing
  - Ever more personalized scams
- Stalking
  - Child predators
- Profiling
  - Identifying compulsive buyers
- Lack of transparency/feedback
  - No idea who knows what about ourselves
  - Information about ourselves revealed by others
Wave imaging (Airport security)

- Advantages [TSA, May 2009]:
  - Effective, fast, convenient, safe
- How privacy is addressed
  - The system uses a pair of security officers
  - A passenger's face is blurred
  - Cameras, cell phones or recording devices not allowed into the room
  - The computers have been programmed so they have “zero storage capability”
  - Images are “automatically deleted”
Wave imaging (Airport security)

What is the problem then?

- Image quality can improve
- Storage capability could be reactivated
- No cameras allowed?
  - Incentives: images of celebrities?
- Policy could be changed down the road
- As machines get cheaper, will we see them in supermarkets or schools?
- General concerns over human rights, personal dignity, proportional?
Smart energy

- Smart meters
  - Better use of resources
  - Send information back to provider (wired/wireless)

- Privacy issues
  - Information that can be inferred
    - At home / away
    - Rutines (e.g., sleep/eat schedule, tv, washing machine…)
    - Changes in routines (e.g., visitors)
The privacy debate

- “If you care so much about your privacy it’s because you have something to hide”
- “Surveillance is good and privacy is bad for national security. We need a tradeoff between privacy and security”
- “People don’t care about privacy”
The privacy debate

“If you care so much about your privacy it’s because you have something to hide”

Solove:

“The problem with the ‘nothing to hide’ argument is its underlying assumption that privacy is about hiding bad things.”
The privacy debate

- “Surveillance is good and privacy is bad for national security. We need a tradeoff between privacy and security”

- “We need more surveillance” is a powerful argument
  - If attacks increase, you can argue that you need even more
  - If attacks decrease, you take credit
The privacy debate

- “Surveillance is good and privacy is bad for national security. We need a *tradeoff* between privacy and security”
- Not **effective**: smart adversaries evade surveillance
- Risk of **abuse**: lack of transparency and safeguards
- Risk of **subversion** for crime/terrorism

Example: Greek Vodafone scandal (2006):
“someone” used the **legal interception** functionalities (backdoors) to monitor 106 key people: Greek PM, ministers, senior military, diplomats, journalists...
The privacy debate

■ “People don’t care about privacy”
■ People want to **control** information:
  - Impression management /self-presentation
    - What do we tell to whom
    - Concerns over information taken out of context
    - We value friends who are discreet
  - Personal safety
■ Lack of transparency / feedback
The privacy debate

- [Solove] “Part of what makes a society a good place in which to live is the extent to which it allows people freedom from the intrusiveness of others. A society without privacy protection would be suffocation.”

- [Diffie and Landau] “Communication is fundamental to our species; private communication is fundamental to both our national security and our democracy.”
Video: ordering pizza
Taking privacy to create security

Is there a tradeoff between privacy and security?

Source: http://www.myconfinedspace.com/
Privacy = Security Property

- **Individuals**
  - Freedom from intrusion, profiling and manipulation, protection against crime / identity theft, flexibility to access and use content and services, control over one’s information

- **Companies**
  - Protection of trade secrets, business strategy, internal operations, access to patents

- **Governments / Military**
  - Protection of national secrets, confidentiality of law enforcement investigations, diplomatic activities, political negotiations

- **Shared infrastructure**
  - Despite varying capabilities infrastructure is shared
  - Telecommunications, operating systems, search engines, on-line shops, software, . . .
  - Denying security to some, means denying it to all: crypto wars redux?
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
Data protection

- 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
  - Unfortunately, millions of records with personal data are breached every year
- Weak enforcement, low penalties
- USA: fair information practices
  - Many individual laws (HIPAA, California disclosure laws)
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
Soft privacy

- System model
  - Data subject provides her data
  - Data controller responsible for its protection

- Threat model
  - External parties, errors, malicious insider
Soft privacy

- Controller: main security “user”
- Policies, access control, trust, audits (liability)
- Goal (data protection): purpose, consent, data security
Soft 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
Soft 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
  - Need to trust data controllers (honesty, competence) and hope for the best
Hard privacy

- System model
  - Subject provides as little data as possible
  - Reduce as much as possible the need to “trust” other entities
- Threat model
  - Adversarial environment: communication provider, data holder
  - Strategic adversary with certain resources motivated to breach privacy (similar to security systems)
Hard privacy

- Subject is an active security “user”
- Goal (data protection): data minimization
- Goal (Solove): protect against surveillance, interrogation, aggregation, identification
- Overview of hard privacy solutions
Some Privacy Enhancing Technologies

- Identity management
- Anonymous e-cash
- Privacy-preserving pay-as-you-drive insurance
- Data anonymization
- Anonymous communications
- Covert communications
- Censorship resistance
Identity Management: partial identities

Legend:
- Identity of Alice
- Partial Identity of Alice

- Government
- Health Care
- Work
- Shopping
- Payment
- Travel
- Leisure
- Telecomcommunication
- Identity of Alice
- Partial Identity of Alice
Entity and attribute authentication

- Entity authentication often first step of a transaction
- Many transactions involve attribute verifications
  - ID documents: state certifies name, birth date, address, …
  - Letters of reference: employer certifies salary for landlord
  - Club membership: the club certifies gold member status
- Credential: token that allows you to certify an attribute
- Entities
  - Issuer (State, Employer, Club)
  - Prover (holder of the credential)
  - Verifier (anyone)
- Property: Prover proves to the Verifier that she holds a credential with certain properties certified by the Issuer
- Properties: Unforgeability and Privacy
Anonymous credentials

- Cryptographic protocols between <Issuer, Prover, Verifier>
  - Prover can prove that he holds a credential with certain attributes
    - or any expression on them (simple arithmetic, boolean) (e.g. salary>30.000 and contract= permanent)
  - Verifier gains no additional information
  - Secure even if Issuer and Verifier collude (single/multiple show)
  - Security: cryptographic (Hard Privacy)
  - Future identity cards and passports?
Anonymous e-cash [Chaum]

- Secure and private payments
  - Cannot forge money or payments
  - With the anonymity of cash
  - Not just cash: public transport
- Anonymous credentials can provide this
  - The bank certifies I have one euro
  - Payment: prover shows the credential, verifier accepts it
  - Verifier goes to the bank to deposit the coin
- Security properties:
  - Unforgeability
  - Privacy (for payer)
  - Double spending prevention!
PriPAYD: car insurance

- GPS + Black box (computation) + transmit billing

- Flexible: easy change
- Easy computation
- Low cost
- Privacy friendly
- Third parties do not carry personal data
Data anonymization

- Anonymized data can be very useful, for example, for research purposes
  - Incidence of diseases: medical research
  - Social network structures: epidemiology, sociology
  - Optimization of services (e.g., transport or computer infrastructures)
- Measure the risk of **re-identification** of anonymized data:
  - Note: data protection does not apply to anonymized data
  - Records in an anonymized database
    - Medical data
    - Internet searches (AOL case)
  - Nodes in an anonymized social graph
Anonymous communications

- Confidentiality of identities (anonymity) and relations (unlinkability):
  - not provided by default by the communication infrastructure
  - The infrastructure is shared by individuals, business, government, military: privacy threats affect all

- Traffic data (origin, destination, time, volume): side channel information
  - Less volume than content: more coarse, but highly valuable information
  - Formats that are easy to process for machines
  - Can be used to select targets for more intensive surveillance
  - Hard to conceal
Steganography and covert communications

- Encryption: hide data content
- Anonymity/unlinkability: hide identities / relations
- **Unobservability**: hide existence
- Communications:
  - Hide the fact that there is any communications
  - Embed a communication within another
  - Covert channels: hide secrets within public information
- Storage:
  - Hide the existence of files
  - Under coercion can deny there are any files to decrypt
Censorship resistance

- How is that privacy technology?
  - Communities, tools, and techniques overlap
  - Second definition: informational self-determination
  - Freedom (and techniques) to communicate, publish or access information

- Censorship resistance in communications:
  - Firewall busting techniques (national firewalls)
  - Peer-to-peer networking and file sharing (combine anon.comms, replicated storage, ...)

- Censorship resistance is the new availability!
Other properties

- **Forward security**
  - Ephemeral keys
  - Minimize consequences of security breach
  - Compulsion

- **Deniability** (*repudiation*)
  - Not possible to prove user knows / has said or done something
  - Communication: off-the-record property
  - Storage: compulsion resistance
Privacy challenges

- Privacy requirements and privacy by design
- Finding efficient and secure mechanisms
  - Complex systems require privacy at every level: the chain is as strong as its weakest link
  - Proposed techniques keep getting broken: lack of models and proofs
  - Secure implementation is even harder
  - Easy to defeat by “changing” abstraction layer
    - camera’s, RFID tags, unique device properties, singulation protocols, traffic analysis, …
New challenging scenarios

- Location privacy
  - Real time
  - Space-Time relation
  - Device fingerprinting
- Ubiquitous environments
  - Constrained devices
  - Securing the physical link
- Social networks
  - Tension with data sharing
- Cloud computing
  - Outsourcing of storage/computations
Conclusions

- Open debate, unclear how it will evolve
- Privacy is not “opposed” to security, but rather a security property
- Compliance is a strong driver
  - Data Protection
  - US disclosure legislation
- Soft Privacy is the state of the art
  - Hidden costs of securing the data silos
- Hard Privacy solutions:
  - Active research
  - Poor deployment: cost/security benefit
- Need for interdisciplinary research
The end

Thank you for your attention