faculty of
Computing &
Digital Technology

Lecture 4

Privacy in Cyberspace

Prepared by:
Ms Anitha Velayutham

Social Change….

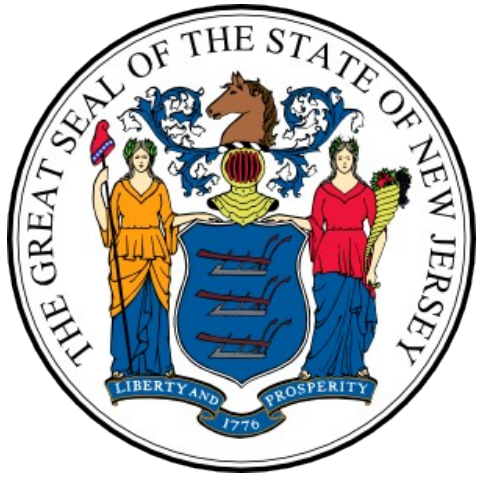
facebook

twitter

Cyber attack's in the news . . .

faculty of Computing & Digital Technology

"Hacked: Data breach costly for Ohio State, victims of compromised info - Breach affects 760,000 people, expected to cost university $4 million"
– The Lantern
December 2010



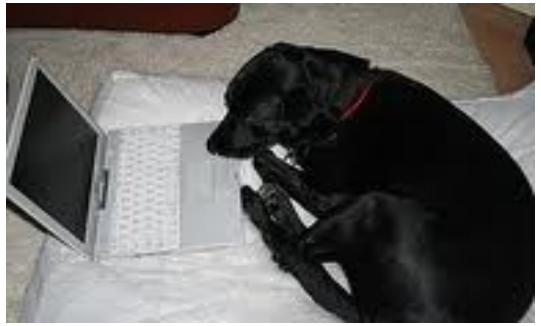
"Hackers broke into the computer system at a New Jersey school district - gained access to student records system used by 160 schools across the state." – *Info Security*
January 2011

"UC-Berkeley records hacked - thieves able to access social security numbers, health files dating back to 1999, over 160,000 records stolen" – *San Jose Mercury News*
May 2009

Protecting Privacy?????

Personal access???

Padlock????



Watch Dog??

Information Technology Erodes Privacy

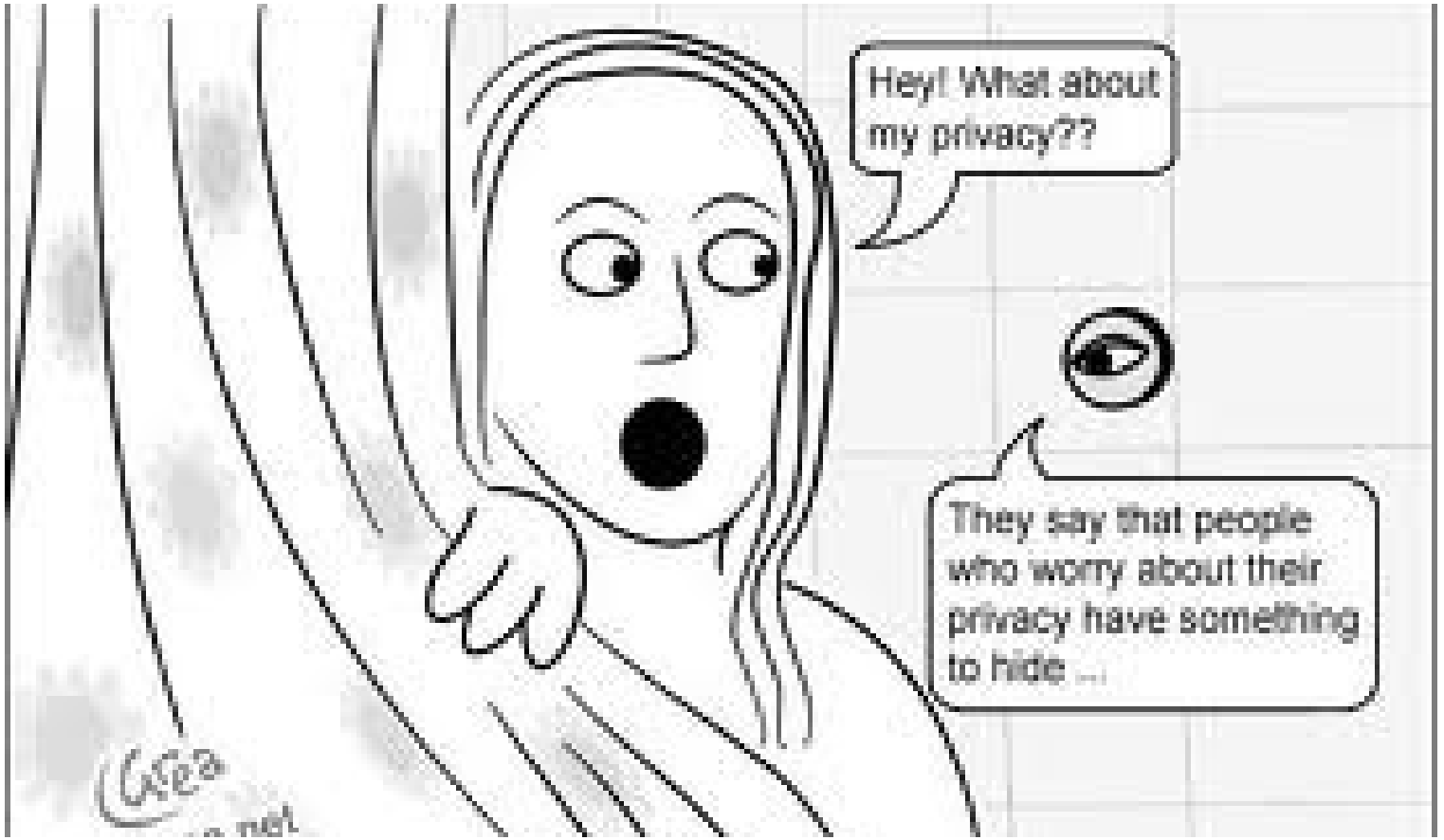
- Computers, databases, and Internet enable ever-improving information __
  - collection
  - exchange
  - combination
  - distribution
- Easier than ever to get information about others, including total strangers
- Scott McNealy: “You have zero privacy anyway. Get over it.”

Is privacy important? If so, can we protect it?

Defining Privacy

- Privacy related to notion of access
- Access
  - Physical proximity to a person
  - Knowledge about a person
- Privacy is a “zone of inaccessibility”
- Privacy violations are an affront to human dignity
- Too much individual privacy can harm society
- Where to draw the line?

Is Privacy important for an individual ???



Privacy as an Important Social Value

- Privacy is important for a diversity of relationships (from intimate to casual).
- It is important for democracy.
- Privacy is an important social, as well as an individual, value.
- Regan (1995) – we need to understand the importance of privacy as a social value.

Three Theories of Privacy

Accessibility Privacy	**Privacy is defined in terms of one's physically "being let alone," or freedom from intrusion into one's physical space.**
Decisional Privacy	**Privacy is defined in terms of freedom from interference in one's choices and decisions.**
Informational Privacy	**Privacy is defined as control over the flow of one's personal information, including the transfer and exchange of that information.**

Is There a Natural Right to Privacy?

faculty of Computing & Digital Technology

Argument in favor

- Right to privacy may have grown out of property rights
  - Europeans have historically viewed the home as a sanctuary
  - English common law tradition: “A man’s home is his castle”
  - Coercive Acts (1773) led to 3rd Amendment to US Constitution: “**No soldier shall, in time of peace be quartered in any house, without the consent of the Owner, nor in time of war, but in a manner to be prescribed by law.”**
- Warren and Brandeis
  - Warren shocked at newspaper coverage of daughter’s wedding
  - “The Right to Privacy” published in 1890
  - Defined privacy as “the right to be let alone”
  - Right to privacy now recognized in courts across America

Is There a Natural Right to Privacy?

Argument against (Judith Jarvis Thomson)

- Nobody seems to know what privacy is
- Problems with defining privacy as “the right to be let alone”
  - One the on hand, definition is too narrow – doesn’t include covert spying
  - On the other hand, definition is too broad – does include assault
- Whenever a right to privacy is violated, another right is violated as well
- Therefore, no need to define privacy or privacy rights precisely

Is There a Natural Right to Privacy?

Conclusion

- Privacy is not a **natural** right, but it is a **prudential** right
- Rational people agree to recognize some privacy rights because granting these rights benefits society

Privacy and Trust

- Perhaps modern life is actually more private than life centuries ago
  - Most people don’t live with extended families
  - Automobile allows us to travel alone
  - Television v. public entertainment
- Challenge: we now live among strangers
- Remedy: establishing reputations
  - Ordeal, such as lie detector test or drug test
  - Credential, such as driver’s license, key, ID card, college degree
- Establishing reputation is done at the cost of reducing privacy

Case Study: New Parents

- Sullivans have a baby girl
- Both work; they are concerned about performance of full-time nanny
- Purchase program that allows monitoring through laptop's camera placed in family room
- They do not inform nanny she is being monitored

Rule Utilitarian Evaluation

- If everyone monitored nannies, it would not remain a secret for long
- Consequences
  - Nannies would be on best behavior in front of camera
  - Might reduce child abuse and parents' peace of mind
  - Would also increase stress and reduce job satisfaction of child care providers
  - Might result in higher turnover rate and less experienced pool of nannies, who would provide lower-quality care
- Harms appear greater than benefits, so we conclude action was wrong

Public Records

- Public record: information about an incident or action reported to a government agency for purpose of informing the public
- Examples: birth certificates, marriage licenses, motor vehicle records, criminal records, deeds to property
- Computerized databases and Internet have made public records much easier to access

The Problem of Protecting Privacy in Public

- *Non-Public Personal Information* (or *NPI*) refers to sensitive information such as in one's financial and medical records.
  - NPI has some legal protection

- Many privacy analysts are now concerned about a different kind of personal information – *Public Personal Information* (or *PPI*).
  - PPI is non-confidential and non-intimate in character – is also being mined.

PPI

- Why should the collection of PPI, which is publicly available information about persons generate controversies involving privacy?
  - For example, suppose learns that that you are a student at Rivier, you frequently attend college basketball games, and you are actively involved in Rivier's computer science club.
  - In one sense, the information is personal because it is about *you* (as a person);but it is also about what you do in the public sphere.

- Hypothetical Scenario:
  - Shopping at Supermart;
  - Shopping at Nile.com;

- Reveal problems of protecting privacy in public in an era of information technology and data mining.

Information Held by Private Organizations

- Credit card purchases
- Purchases made with loyalty cards
- Voluntary disclosures
- Posts to social network sites

Three Ways Privacy is Threatened by Cybertechnology?

- *data-gathering* techniques used to collect and record personal information, often without the knowledge and consent of users.
- *data-exchanging* techniques used to transfer and exchange personal data across and between computer databases, typically without the knowledge and consent of users.
- *data-mining* techniques used to search for patterns implicit in large databases in order to generate consumer profiles based on behavioral patterns discovered in certain groups.

Computerized Merging and Matching Operations

- *Computer merging* is a technique of extracting information from two or more unrelated databases, which contain data about some individual or group of individuals, and incorporating it into a composite file.

- *Computer matching* is a technique that involves the cross checking of information in two or more databases that are typically unrelated in order to produces certain "matching records" or "hits."

Data Gathering and Privacy Implications

- Facebook tags
- Enhanced 911 services
- Rewards or loyalty programs
- Body scanners
- RFID tags
- Implanted chips
- Mobile apps
- Facebook Login
- OnStar
- Automobile “black boxes”
- Medical records
- Digital video recorders
- Cookies

Facebook Tags

- Tag: Label identifying a person in a photo
- Facebook allows users to tag people who are on their list of friends
- About 100 million tags added per day in Facebook
- Facebook uses facial recognition to suggest name of friend appearing in photo
- Does this feature increase risk of improper tagging?

R F I D Tags

- RFID: Radio frequency identification
- An RFID tag is a tiny wireless transmitter
- Manufacturers are replacing bar codes with RFID tags
  - Contain more information
  - Can be scanned more easily
- If tag cannot be removed or disabled, it becomes a tracking device

R F I D Tags Speed Inventory Process

Employees take inventory more quickly and make fewer errors when items are marked with RFID tags. (Marc F. Henning/Alamy)

Implanted Chips

- Taiwan: Every domesticated dog must have an implanted microchip
  - Size of a grain of rice; implanted into ear
  - Chip contains name, address of owner
  - Allows lost dogs to be returned to owners

- RFID tags approved for use in humans
  - Can be used to store medical information
  - Can be used as a “debit card”

Mobile Apps

- Many apps on Android smartphones and iPhones collect location information and sell it to advertisers and data brokers
  - Angry Birds
  - Brightest Flashlight
- Flurry: a company specializing in analyzing data collected from mobile apps
  - Has access to data from > 500,000 apps

Facebook Login

- Allows people to login to Web sites or apps using their Facebook credentials
- App's developer has permission to access information from person's Facebook profile: name, location, email address, and friends list

Automobile “Black Boxes”

- Modern automobiles come equipped with a “black box”
- Maintains data for five seconds:
  - Speed of car
  - Amount of pressure being put on brake pedal
  - Seat belt status
- After an accident, investigators can retrieve and gather information from “black box”

Medical Records

- Advantages of changing from paper-based to electronic medical records
- Quicker and cheaper for information to be shared among caregivers
  - Lower medical costs
  - Improve quality of medical care
- Once information in a database, more difficult to control how it is disseminated

Digital Video Recorders

- TiVo service allows subscribers to record programs and watch them later
- TiVo collects detailed information about viewing habits of its subscribers
- Data collected second by second, making it valuable to advertisers and others interested in knowing viewing habits

Cookies

- Cookie: File placed on computer's hard drive by a Web server
- Contains information about visits to a Web site
- Allows Web sites to provide personalized services
- Put on hard drive without user's permission
- You can set Web browser to alert you to new cookies or to block cookies entirely

General Data Protection Regulation

- General Data Protection Regulation (GDPR): set of rules governing collection of information from citizens of European Union
- Requires companies to…
  - Disclose information they are seeking to collect
  - Disclose why they are collecting it
  - Get permission before collecting it
- Responding to GDPR, most large American companies are adopting new privacy guidelines
  - Web-site banners informing users, asking for consent

SCHOOL OF ICT

faculty of
Computing &
Digital Technology

DATA MINING

Data Mining Defined

- Data mining involves the indirect gathering of personal information through an analysis of implicit patterns discoverable in data.
- Searching records in one or more databases, looking for patterns or relationships
- Can be used to create profiles of individuals
- Allows companies to build more personal relationships with customers

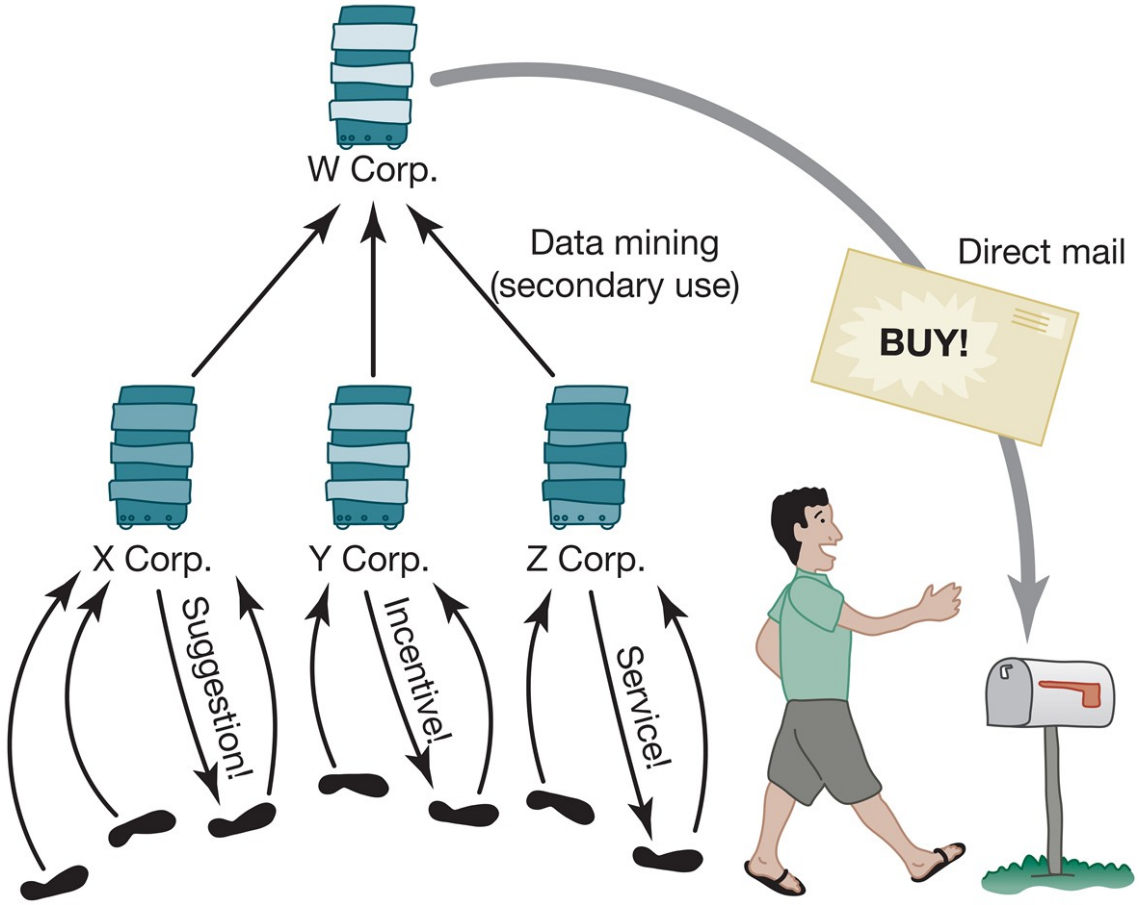
Google's Personalized Search

- Secondary use: Information collected for one purpose use for another purpose
- Google keeps track of your search queries and Web pages you have visited
  - It uses this information to infer your interests and determine which pages to return
  - Example: "bass" could refer to fishing or music
- Also used by retailers for direct marketing

Limiting Information Google Saves

- You can limit amount of information Google saves about your activities
- Privacy Checkup lets you pause collection of personal information
  - Search queries and other Google activity
  - Location information collected from signed-in devices
    - Where you have gone
    - How often you have gone there
    - How long you have stayed
    - Customary routes of travel
  - Contact and calendar information
  - Recordings of your voice and accompanying audio
  - YouTube search queries
  - YouTube videos you have watched

Secondary Uses of Information



Collaborative Filtering

- Form of data mining
- Analyze information about preferences of large number of people to predict what one person may prefer
  - Explicit method: ask people to rank preferences
  - Implicit method: keep track of purchases
- Used by online retailers and movie sites

Ownership of Transaction Information

- Who controls transaction information?
  - Buyer?
  - Seller?
  - Both?
- Opt-in: Consumer must explicitly give permission before the organization can share info
- Opt-out: Organization can share info until consumer explicitly forbid it

Credit Reports

- Example of how information about customers can itself become a commodity
- Credit bureaus
  - Keep track of an individual's assets, debts, and history of paying bills and repaying loans
  - Sell credit reports to banks, credit card companies, and other potential lenders
- System gives you more choices in where to borrow money
- Poor credit can hurt employment prospects

Targeted Direct Mail

- Businesses mail advertisements only to those most likely to purchase products
- Data brokers provide customized mailing lists created for information gathered online and offline
- Example of making inferences for targeted direct mail
  - Shopping for clothes online + frequent fast-food dining + subscribing to premium cable T V channels → more likely to be obese
- Two shoppers visiting same site may pay different prices based on inferences about their relative affluence

Social Network Analysis

- Collect information from social networks to inform decisions
- Bharti Airtel (India) offers special promotions to “influencers”
- Police use Facebook and Twitter posts to deploy officers on big party nights
- Banks combine social network data with credit reports to determine creditworthiness

Netflix Prize

- Netflix offered $1 million prize to any group that could come up with a significantly better algorithm for predicting user ratings (2006)
- Released more than 100 million movie ratings from a half million customers
  - Stripped ratings of private information
- Researchers demonstrated that ratings not truly anonymous if a little more information from individuals was available
- U.S. Federal Trade Commission complaint and lawsuit
- Netflix canceled sequel to Netflix Prize (2010)

AOL Search Dataset

- AOL researcher Dr. Chowdhury posted three months' worth of user queries from 650,000 users (2006)
- No names used; random integers used to label all queries from particular users
- Researchers identified some users from queries; e.g., many people performed searches on their own names
- New York Times investigation led to public outcry
- AOL took down dataset, but already copied and reposted
- AOL fired Dr. Chowdhury and his supervisor

Examples of Consumer or Political Backlash

Marketplace: Households

- Lotus Development Corporation developed CD with information on 120 million Americans
- Planned to sell CD to small businesses that wanted to create mailing lists based on various criteria, such as household income
- More than 30,000 consumers complained to Lotus about invasion of privacy
- Lotus dropped plans to sell CD

Facebook Beacon

- 2007: Facebook announced Beacon, a targeted advertising device
  - Facebook user makes purchase
  - Facebook broadcasts purchase to user's friends
  - Based on opt-out policy: users enrolled unless explicitly asked to be excluded
- A significant source of advertising revenue for Facebook
- MoveOn.org led online campaign lobbying Facebook to switch to an opt-in policy
- Mark Zuckerberg apologized, and Facebook switched to an opt-in policy

Malls Track Shoppers' Cell Phones

- In 2011 two malls recorded movement of shopper by tracking locations of cell phones
  - How much time people spend in each store?
  - Do people who shop at X also shop at Y?
  - Are there unpopular areas of mall?
- Small signs informed shoppers of study
- After protest, mall quickly halted study

iPhone Apps Upload Address Books

- In 2012 a programmer discovered Path was uploading iPhone address books without permission
- Internet community pointed out this practice violated Apple's guidelines
- CEO of Path apologized; app rewritten
- Twitter, Foursquare, and Instagram also implicated for same practice

Cambridge Analytica

- Robert Mercer's vision: Use data analytics to help conservative candidates and causes
- Mercer formed joint venture with SCL Group and invested $15 million in new firm: Cambridge Analytica
- SCL Group hired Aleksandr Kogan to gather data about American voters
- Kogan created survey app: "thisisyourdigitallife"
  - Promoted survey using Amazon's Mechanical Turk
  - Users paid $1 or $2 to take personality test
  - Users had to access app using Facebook Login
  - Users agreed that app would download information about them **and their Facebook friends22**

Cambridge Analytica

- Personal data collected from 270,000 people who took surveys and as many as 87 million people who were on their friends' lists
- Kogan shared profiles with Cambridge Analytica
- About 30 million profiles were detailed enough that Cambridge Analytica could combine data with other data they had, creating psychographic profiles
  - Classified voters over five personality traits: openness, conscientiousness, extroversion, agreeableness, neuroticism
  - Strategy: target ads based on psychographic profile
- Ted Cruz campaign hired Cambridge Analytica to help with microtargeting
  - Value of advice debatable
  - Campaign staffers said predictions were bad

Cambridge Analytica

- Trump campaign hired Cambridge Analytica in fall 2016 – firm promised to provide names of millions of voters likely to vote for Trump
- “Data breach” story broke in spring 2018
  - Facebook response
    - Not a breach – everyone who used Kogan’s app had granted their consent, and privacy settings of their friends allowed their information to be shared
    - Kogan had perpetrated a fraud by sharing data with Cambridge Analytica
    - Suspended accounts of Kogan and Cambridge Analytica
  - Mark Zuckerberg called to Washington, D C, and testified for 10 hours in front of two Congressional Committees
- May 2018: Cambridge Analytica filed for bankruptcy

Comprehensive Privacy Proposals

- Clark argues for a "co-regulatory" model.
- He believes that a successful on-line-privacy policy must include:
  - strong legislation;
  - a privacy oversight commission;
  - industry self-regulation.
- These must also be accompanied by privacy-enhancing technologies.
- A "privacy watchdog agency" and sanctions are also both needed.

Summary

- Modern information technology makes it much easier to collect and transmit information
- Privacy a balancing act
  - Desires of individuals
  - Profit motives of companies
  - Common good
- Public records: information that communities have decided should be known to all
- Sometimes must share personal information to get something we want
  - Disclose income tax statements to get a home loan
- Companies collect more information to market more selectively – some have pushed the boundaries of what society will tolerate

"Instead of waiting for someone to steal my identity, I'm going to auction it on eBay!"

Reference

- Ethics for the Information Age, 8th Edition by Michael J. Quinn (Pearson)
- Ethics and Technology: Ethical Issues in an Age of Information and Communication Technology by Tavani H. T.