Lecture #22: Privacy

CS106E, Young

This lecture we look at privacy issues. We begin by considering how computing technology has affected our personal privacy. As we discover, our movements and actions can be tracked much more easily in today's society, and technology allows processing of that information much more easily as well. We consider laws related to privacy in the US and in Europe.

Consumer privacy advocates sometimes point out that when using a service you are either a "customer or a product". We consider what that means and whether or not we should be concerned with use of our data by commercial entities. As we discover, even if the commercial entity we are working with has the greatest intention to use our data in a reasonable manner, if their computer security is poor, that data may end up in public anyway.

We look at tools and techniques used to attack or defend privacy including web beacons or tracking pixels, third party cookies, email tracking, and hiding one's tracks online using the TOR network.

We consider whether or not anonymity online is a good or bad thing, then look at some ways in which governments might use that information. China provides an example of how a government can use the power of computer technology to monitor and potentially modify citizens' behavior. We look at their proposed Social Credit System, which aims to take all information on citizens and give them a score. We look at how pilot projects have used those scores to control access to government services or even to help with match making on dating websites.

The ability of modern computers to process Big Data has a considerable impact on our privacy. We end the lecture by taking a look at issues of big data including considering the 3Vs of Big Data – the Volume, Velocity, and Variety of data available. We'll look at artificial intelligence and machine learning techniques including those related to big data in the next lecture.

Why Privacy is Different in the Digital Age

Privacy is particularly problematic given our reliance on technology. Let's take a look at a few of the reasons:

Information Available for Tracking – We are constantly leaving an electronic trail which can be used to track our movements and our actions.

- Cell Phones can be used to track our movement.
- While Cell Phone towers only provide a general area, WiFi tracking provides even finer grain tracking.

    - For example, Nordstrom’s has experimented with using customer’s Cell Phone WiFi connections to track movement between departments in their stores. This is much finer grain tracking than cell phone towers can provide.
  - Closed Circuit Cameras can track where we go and what we do.
    - The United Kingdom is estimated to have one closed-circuit television camera for every 10-15 people.
  - As we consume most of our media electronically or directly online, our reading and viewing habits are now very easy to track. The government or a corporation can review which ebooks we read, which newspaper sites we visit, what music we listen to, and which movies and tv shows we watch.

Computers provide Analysis – Computers allow us to process data in ways not previously available. License plate reading can be combined with traffic cameras to track our vehicles. Face recognition combined with closed-circuit television (CCTV) cameras makes it easy to track our movements.

Big Data – The ability to store massive amounts of information and the ability to combine data from many different sources gives both commercial and government users capabilities previously unseen.

Legal Issues

US Law

- The US Supreme Court has ruled that individuals have a right to *Practical Obscurity*.
- Using this doctrine, requests for access to FBI rap sheets using the Freedom of Information Act (FOIA) were turned down, even though all information on the sheets were publically available.
- Gathering the information from various public sources was considered a difficult task, and those listed on the rap sheets had a right to the practical obscurity granted by the difficulty of gathering all the publically available information.

- In the era of big data and the Internet, the Practical Obscurity doctrine is under assault, as companies can easily gather the data and sell it to the public.

European Right to be Forgotten

- In 2014, the European Court of Justice ruled that in certain circumstances citizens had a right to be forgotten.
- Under this ruling, Search Engines had to remove certain Search Results.
  - Newspapers did not necessarily have to remove articles related to the citizens, but search engines could not link to them.
  - Within half a year of the court ruling, Google had received over 120,000 requests for deletion. (source)

European General Data Protection Regulation (GDPR)

- In 2016 the European Union agreed to the GDPR. It became active May 25, 2018.
- This regulation includes:

Breach Notification – Individuals must be notified of a security breach of their information within 72-hours of the breach.

Right to Access – Individuals have the right to access and view information a company stores about them.

Right to be Forgotten – Individuals have the right to have data removed.

Privacy by Design – Companies are reminded that going forward they should have privacy designed in to their products from the start.

Data Protection Officers – Companies falling into certain categories (those "whose core activities consist of processing operations which require regular and systematic monitoring of data subjects on a large scale") must have Data Protection Officers.

- Violations of the GDPR may result in penalties up to 4% of a company's annual worldwide revenue.

Commercial and Government Issues
Privacy Issues arise from both commercial and government use of information. We'll start by looking primarily at commercial concerns and then will take a look at governmental concerns.

Product or Customer

- Privacy advocates sometimes point out that when you're using a commercial service, you are either the customer or the product.
  - This is the tech equivalent of "there's no such thing as a free lunch."
  - While some may argue that this oversimplifies the situation, we can certainly agree that companies must make money somehow.
    - If you're using a service and not paying for it, one common approach is for the company to sell your information either directly or via targeted advertising.
- Unfortunately, you can be a paying customer and still have your data sold for further profit.
  - As of April 2017, Internet Service Providers can now sell information on your browsing habits to others.

Should we be concerned with Commercial use of Our Information?

- Perhaps I'm perfectly happy to use Google for free, and I prefer to see advertisements from the DoubleClick advertising network for products I might be interested in over advertisements for things I have no interest in. Should I still care about commercial use of my information?
- One classic example of abuse of information is, what happens if you visit websites related to cancer. Would you be concerned if your ISP passed that information on to your insurance company and they raised your insurance rates?

Legal and Illegal Access to Information

- When you provide information online, keep in mind that your information may be accessed legally or it may be accessed illegally. It is only as safe as the computer security of whoever is holding it.
- Let's take a look at a few example:

Equifax Security Breach

- In 2017 the Equifax credit service's computers were breached.
- ~150 million people's data was exposed
- data taken included names, birthdates, and social security numbers
- in some cases, driver's license numbers were taken
- a credit card can be issued if a scammer has a person's name, social security number, and a driver's license number

- The people whose data was taken aren't customers of Equifax, even though Equifax had their information, so they don't even have the recourse of taking their business to someone else.

Ashley Madison

- In 2015 the Ashley Madison website for adulterous affairs was hacked.
- 30 million names of users on the website were leaked.

- Even if you trust a business not to expose your information directly, do you trust their computer security experts enough to be sure that hackers won't get your information?

Facebook

- Data from over 50 million profiles was taken and then sold to Cambridge Analytica.
    - Facebook users were falsely told by a Cambridge professor that their data would only be used for academic purposes.
    - Data taken included not only those who agreed to participate in a survey, but those of their friends as well.
    - The Cambridge professor was dishonest and violated Facebook's rules. However, nevertheless the data was exposed.
- Facebook has fired several employees for abusing their position to cyberstalk others.

Tools and Techniques

Web Beacons/Bugs

- Web Beacons or Bugs are used to track users on the Internet.
- One common method for doing this is to include an invisible 1x1 pixel image within an HTML file.
    - This image, sometimes referred to as a *tracking pixel*, has a unique name, which includes an ID number in it that can be used to identify a specific visitor.

Email Tracking

- Use of invisible images can also be used in conjunction with emails.
    - An HTML formatted email is sent including a tracking pixel.
    - When the user views the email, an HTTP request is sent to the web server.
    - By tracking HTTP requests for specific image files, an email sender can see exactly when an email is viewed.
- This technology can be used to determine when a business partner reads an email. It can also be used by Spammers to determine whether an email address is active or not.
- Many email programs and email websites provide the option of not downloading images within email messages to prevent email tracking via web beacons.

Tracking via Third Party Cookies

- A website can allow a third party such as an advertising network to track a user's actions across multiple websites.
- If a website includes either an invisible web beacon or a visible advertisement from a third party sever (a server that is in addition to the main web server the user is accessing)[1] that website can place a cookie with a unique identifier on the user's web browser.
- When the user visits another website that also uses that same advertising network, the HTML file contains a reference to the third party web server (again using a visible advertisement or an invisible tracking beacon). When the HTTP request for that image is sent, the cookie information with the user's unique identifier is also sent.
- If the user enters their name, address, or phone number on any website within the advertising network, this can be passed on to the advertising network and combined with the user's unique identifier to create a non-anonymous picture of a specific individual and their web browsing habits.

[1] The first party is the user, the second party is the website they are visiting, and the third party is the advertising network.

TOR Anonymous Routing

- TOR is a system that allows users to remain anonymous as they carry out actions on the Internet.
- You can think of TOR as a super-powered VPN, as it provides similar function, but provides even better protections at a cost in speed.
- TOR stands for "The Onion Router" and refers to the way the TOR network is multi-layered like an onion.

- When a user accessing a server via TOR, their information is encrypted and passed through a series of different servers on the TOR network.
  - As of I'm writing this, there are over 6000 servers currently running on the TOR network.
- A TOR server routing sequence is only good for 10 minutes, further accesses will use a different random sequence of servers to reach their destination.

- While intermediate servers cannot see the data within TOR traffic, whichever server last handles the data before it is passed back to the regular Internet can read the data. Therefore, TOR users should use their own encryption when communicating via TOR (e.g., use HTTPS when communicating with a website or secure messaging when communicating with someone else).

Should we be concerned with Government use of Our Information?

- I've joked in class that I don't really care if the NSA can read my emails. I don't have anything they will be concerned with. However, there are wider issues at stake. Consider the following examples:

  - Online anonymity helps drug traffickers and terrorist organizations.
  - Anonymity allows online harassment of individuals often with no recourse available.

  - Anonymity allows women's rights, gay rights, and human rights activists to operate in countries where they might face harassment or worse from the government.

Chinese Government Use of Computers

- The Chinese Government probably provides us with our best look at how a government can use computer surveillance to monitor and control their citizens.

Social Credit System

- The Chinese Government is creating a ***Social Credit System***.
- According to the Washington Post:

  … the Communist Party hopes will build a culture of "sincerity" and a "harmonious socialist society" where "keeping trust is glorious."

  The ambition is to collect every scrap of information available online about China's companies and citizens in a single place — and then assign each of them a score based on their political, commercial, social and legal "credit."

- Here's how The Atlantic describes the system:

  … the system provides abundantly for sticks as well as carrots. Attend a "subversive" political meeting or religious service, for example, or frequent known haunts of vice, or do under-the-table business with an unregistered, informal enterprise, and the idea is that

the network will know about it and respond by curtailing one's privileges. The state wants its citizens to believe that there's little point in trying to evade detection of such acts...

- According to the BBC, eight Chinese companies are experimenting with state-sponsored social credit pilot programs.
  - One program's social credit score is used by Baihe, China's biggest matchmaking service.

- In one program in Suining Province, citizens were given ratings from A-D.

  citizens were classified into four levels: Those given an "A" grade qualified for government support when starting a business and preferential treatment when applying to join the party, government or army; or applying for a promotion.

  People with "D" grades were excluded from official support or employment.

CCTV Cameras, Face Recognition, and Minority Ethnic Groups

- The Social Credit System can be combined with CCTV cameras and face recognition systems.
- The New York Times reports that the Chinese government has used Facial Recognition technology to "track and control Uighurs, a largely Muslim minority".
  - The system "looks exclusively for Uighurs based on their appearance and keeps records of their comings and goings for search and review."
  - Moreover, China is exporting its surveillance technology to other countries.

Installation of Full Spyware on Cell Phones

- The Washington Post reported that an app published by China's Communist Party includes a backdoor giving an external user full "superuser" access to the phone including "allowing the app to access and take photos and videos, transmit the user's location, activate audio recording, dial phone numbers and trawl through the user's contacts and Internet activity, as well as retrieve information from 960 other applications including shopping, travel and messaging platforms."
- The app "collects and sends detailed log reports on a daily basis, containing a wealth of user data and app activity."
- All members of the Communist party were directed to download the app. Other companies and organizations including "Beijing Chaoyang Lawyer's Association and Peking University" have also ordered their members to download the app.

Big Data

- Combining our ability to track our movements, record our purchases, monitor our conversations, and see what we read and watch with a computer's analytical capability could lead to a revolution or nightmare in how our society works.
- Big Data refers to our ability to process all this data.

- Improvements in computing technology have made Big Data analysis possible. These include:
  - Storage costs have dropped, allowing storage of a great deal of information.
  - Computing costs have dropped giving us more processing power.
  - Techniques such as machine learning have given us the ability to analyze large amounts of data.

The 3Vs (or the 4Vs) of Big Data

- Big Data experts refer to the 3Vs which can be used to characterize Big Data (some experts have added a fourth V). These are:

Volume – The amount of data.

Velocity – The rate at which data is generated. Depending on the velocity, different techniques may be used for processing. Information may be batch processed (gathered together and then processed all at once) or stream processed (handled as it comes in). Depending on the velocity of data, we may or may not be able to handle the data in real-time.

Variety – Different types of data exists. Data may be structured or unstructured. Structured data is well suited for computer processing. Unstructured data is not. For example, a passenger manifest is structured data. A CCTV camera feed of people arriving and departing from a passenger terminal is unstructured.

Veracity – How accurate is our data. How certain are we of our data sources.

Target Stores Pregnancy Detection – A Big Data Case Study

- Target stores set out to identify expecting mothers. They wanted to identify them as early as possible. Using public birth records was something their competitors could do just as easily, so they turned to Big Data analysis for something more sophisticated.
- Target, as with many companies has discount club cards.
    - These cards provide discounts on some items and in turn can be used to track the actions of specific customers.
    - Records can be built including such data as:
        - age, marital status, number of kids
        - distance to the nearest Target store
        - estimated salary
        - whether customer has recently moved

- Target had some other data to work with, as they have a Baby Shower Registry.
- Target discovered that
    - Pregnant women purchased large quantities of unscented lotion at the beginning of the 2nd trimester.
    - They purchased supplements of calcium, magnesium, and zinc for the first 20 weeks.
    - They purchased scent-free soap and extra-large bags of cotton balls.

    - In total, Target identified 25 items that could act as pregnancy predictors.