CS573 Data Privacy and Security

Li Xiong

Department of Mathematics and Computer Science Emory University

Today

- Meet everybody in class
- Course overview
- Course logistics
- Poll

Instructor

- Instructor: Li Xiong
 - Web: http://www.mathcs.emory.edu/~lxiong
 - Email: <u>lxiong@emory.edu</u>
 - Office Hours: TuTh 5:15-6:15pm
 - Office: MSC E412

About Me

- Graduate teaching
 - CS550 Database systems
 - CS570 Data mining
 - CS573 Data privacy and security
- Research
 - data privacy and security
 - information integration and informatics

Meet everyone in class

- Group introduction (2-3 people)
- Introducing your group
 - Names
 - Your goals for the course
 - Something interesting about your group

Today

- Meet everybody in class
- Course overview
- Course logistics
- Poll

What is the course about

- Techniques for data privacy and security
- Applications

- Not about
 - Network security, system security, software security ...

Definitions of Privacy

- Right to be left alone (1890s, Brandeis, future US Supreme Court Justice)
- a: The quality or state of being apart from company or observation; b: freedom from unauthorized intrusion (Merrian-Webster)
- The right of individual to be protected against intrusion into his personal life or affairs, or those of his family, by direct physical or by publication of information (Calcutt committee, UK)

Aspects of Privacy

Information privacy

Collection and handling of personal data, e.g. medical records

Bodily privacy

- Protection of physical selves against invasive procedures, e.g. genetic test
- Privacy of communications
 - Mail, telephones, emails
- Territorial privacy
 - Limits on intrusion into domestic environments,
 e.g. video surveillance

Information Privacy

- Establishment of rules governing the collection and handling of personal data
 - Data about individuals should not be automatically available to other individuals and organizations
 - The individual must be able to exercise a substantial degree of control over that data and its use.

Models of privacy protection

- Comprehensive laws
 - Adopted by European Union, Canada, Australia
- Sectoral laws
 - Adopted by US
 - Financial privacy, protected health information
 - Lack of legal protections for data privacy on the Internet
- Self-regulation
 - Companies and industry bodies establish codes of practice
- Technologies of Privacy

A race to the bottom: privacy ranking of Internet service companies

- A study done by Privacy International into the privacy practices of key Internet based companies in 2007
- Amazon, AOL, Apple, BBC, eBay, Facebook, Google, LinkedIn, LiveJournal, Microsoft, MySpace, Skype, Wikipedia, LiveSpace, Yahoo!, YouTube

A Race to the Bottom: Methodologies

- Corporate administrative details
- Data collection and processing
- Data retention
- Openness and transparency
- Customer and user control
- Privacy enhancing innovations and privacy invasive innovations

A race to the bottom: interim results revealed

Privacy-friendly and privacy enhancing

Generally privacy-aware but in need of improvement

Generally aware of privacy rights, but demonstrate some notable lapses

Serious lapses in privacy practices

Substantial and comprehensive privacy threats

Comprehensive consumer surveillance & entrenched hostility to privacy

		CONTROL LUNGUISHER	Contracting and	Outo Received	Transparated	Participation of the same of t	EURCHI COMPANI	Customer Control	Fair Gatesyaya	Extraox.Futurestation	A. C.	Austria
		Salvanianis print Alband Logis correspondent to the salvanian service of the salvanian service o	Section 2011 of the Control of the C	Mary methods and the remaining	Protect meeting in Information Street Pro- Information is read to		Standard philosophic in manner published begins for straightful of the Park appearing straightful one. Park pages of dearther problems on	mainterrunt is ready obtained management tree obtain principles by tree obtain formall many to Arreston.	Company (Company) of the Company of			decimalism being bringers critical and resident and the control critical and resident and the control designation and the control and production from the control and production of the control and
			Management of the Community of the Commu	discontrational reconstitution	STREET, ST. STREET, ST	PROTECTION OF THE PROT	STANDARD STANDARD STANDARD STANDARDS	Colorating Communities in Society of Colorating Spring From African Colorating Springs	Country of the Countr	Fig. 69 Feeder Additions opposed the critical formation and the critical fo		Marriage D. Proposition St. Ac.
-		Particular Property and Control of the Control of t	derjouwhentstell, strike och tennes och tenn	Spin, appoint Photology, e.g. Man Spin, appoint Photology, e.g. Man Spin, e.g. Man Spin, e.g. Man Spin, e.g. Ma	School of Hilliam Andread Countries Countries and Countries Countr	morph resistant to the street of the street	Problementalis - Section Co. S	Print make our or house Still delice to make TVL John C	STANTON PROMINETARIO OF STAN SAN ORGANIZATION OF STAN SAN ORGANIZATION OF STANDARD	Facilities adjust not retained for management of the control of th		
	BEV.GETYWE.		Control of the contro	Estimaterosac	Patrician policy in about	TRANSPORTED TO		Explains how to opt-out		No. 100 Property and Property and Pro-	Charles and the same of	From the Above of States o
	BEET TO PETER	Continue patient goods Create Postson on 17th, affice management	The state of the s	Management and August Proposed Page	Become Change Chan	Senter Windows and Senter Sent		ENLAND OF STREET	Service and the service and th	Story West Con-	PARISHED ENGINEERS	Estat and belong the
		Corporate Leadership		Onto Recentron	Principian and	**************************************	server company	Sustamor Control	Fair Gateways	Section of the section of		200000000000000000000000000000000000000
-	ELECTRICATE STATE	Manufact of Yeart or	Intermeter competee	troundere	Decomposition appearance	STATE OF THE PROPERTY OF		STREET, STREET	was and access to	A tot of the content or order of the content or the content of the content or the content of the content or the		Character Section Cont.
	AND MULTIPLE LOWERS	Electronical of Tribula of	Control of the contro	the automorphism resulting	Shareto perferency produces		Parametria to force toop principles - 6 years be- principled - 6 years	Legislation for Pully capit condi- grand and condition of the Con- pleted and condition of the Con- pleted and condition of the Con- pleted and Condition of the Con- lete and Condition of the Condition of the Condition of the Con- lete and Condition of the Condit	To Book a secretary of the secretary to	Mountheacon-to-		Figure 1 and the state of the s
	Pers Samuel Pers at Fraginal and Samuel Persons and		And the second s	Manufacture country	SERVICE CONTROL			Literate Printing officialists for the community of the printing of the community of the printing of the community of the designation of the community of the community of the printing of the community of the printing of printing of the printing of printing of printing printing printing printing printing printing printing prin	PARAMETER CONTRACTOR PROPERTY OF THE PARAMETER OF THE PAR	Research Action		
	STATE OF THE RESIDENCE	Fingleschand accounts to discharge the state of the state	Establishment of the Control of the	XXXXXXXXXXX	processor and the same	Acceptance of American Matter to	mid-tomory and page to the property of the page of the	Constitution to the constitution of the consti	CONTRACTOR ACCORDANCE FOR CONTRACTOR AND CONTRACTOR ACCORDANCE AND CON	Marie colling Franchischer et eleganting for the party of the colling of the coll		
	Burney Smarth.			No. Hilliam and a constitution of the constitu	SEAST SEATURE OF COLUMN			March 1980, Mr. C. W. C. Salari March 1980, J. Silver, J. L.		Mo. Hillian and the constitution of the consti		Propagation of the Control of the Co
	- Control of the Cont	Corporate Landsontile		Onto Returnitors	Westerman and	Responsitues	Extract Company	Contours Control	Fate distances	Private Phototop of	A. Postini	Anna Contract Contrac
	STATE OF A STATE OF S		The second state of the se	tio_nature	Zine.			Children to Francis against britados.	Section of the sectio	depression of goldens, course		In the second se
	SECTION SECTION	Market Plantage Vision and	Company of the second s	Many change according actor	Princed's positive configurate manner agreement form of the con- constitution and the con- constitution and the con-			discovered between and tensor the second of tensor	Liberton Matthew Elections and the Control of the	CATON CONTROL OF THE PARTY OF T		
	25.		Chapter Whom Service more between the control of th	AMENING ASSESSMENT OFFICE ASSESSMENT OF THE PROPERTY OF THE PR	March C. M. Ward			possible change in		A.A. The second		
	HELLER FLOTON	Control of	TOTAL CONTROL OF THE STATE OF T	Salara dan Paramenta and Research Committee of the Commit	PERSONAL SERVICE AND ADDRESS OF THE PERSONAL PROPERTY OF THE PERSONAL P	Transport Transport Constitution of the second C	Company of the design	Participation of the same	At the Change of the Control of the	SHIPP TO STATE OF THE STATE OF		
		moreover bendereben	to the second second	Mate Metaclico	VERRIERARIO	Bosponstvennes	STREET PROPERTY	PRODUCT POSTOR	Patr Material	EXCESSES PRODUCES AND	A-2000000	Part of The set Inc
	SAND TOWNS SERVICE.			ENGLISHED CONTRACTOR		COMMISSION STATES		(CONTRACTOR		Management -		
	MANAGETA			Description of the Control of the Co	Mark Jerosan and San			Acceptance out a stream	PAGES JAMES	Mountaine		
			70.0F	TOOL WATER TO SHARE THE PARTY OF THE PARTY O	The state of the s		Established		Dat government		~~**********	pattie dissipa
	Freedometopens				Management of the same					The same of the sa		SCHOOLSK-SETTERAL
	\$325035000000000000000000000000000000000		ECCASION AND AND AND AND AND AND AND AND AND AN		Electric Action of the Control of th	Bination others, he - a - version and			Date of the second			
	ESPECIAL PROPERTY OF THE PROPE	more was one on disc.	The second second	ato-seasif-E-Tourseason	Mess recommendation of content.			hand-hand-drawl-seven-	Sequences consequences	Introduction Commence		Manager Street Street
	months of the second		200	Managarat Grandelina	Energy and a second of the		NO. NO. AND ADDRESS OF THE PARTY OF THE PART		lantic towardon and provide	DESIGNATION OF THE PERSON OF T		
					Transmin and					ENGINEER WOODS ON		
	EMMES SEASONS CO.			and the same of the same of	- Charles to come and come	- 5000-00-00-00-00-0			33000000000000000000000000000000000000	Constitution of the consti		
				SPERMSPECKER***	WESTERS C.	-			Manufacture and a second as a		-	

A race to the bottom: interim results revealed

Privacy-friendly and privacy enhancing

Generally privacy-aware but in need of improvement

Generally aware of privacy rights, but demonstrate some notable lapses

Serious lapses in privacy practices

Substantial and comprehensive privacy threats

Comprehensive consumer surveillance & entrenched hostility to privacy



Why Google

- Retains a large quantity of information about users, often for an unstated or indefinite length of time, without clear limitation on subsequent use or disclosure
- Maintains records of all search strings with associated IP and time stamps for at least 18-24 months
- Additional personal information from user profiles in Orkut
- Use advanced profiling system for ads

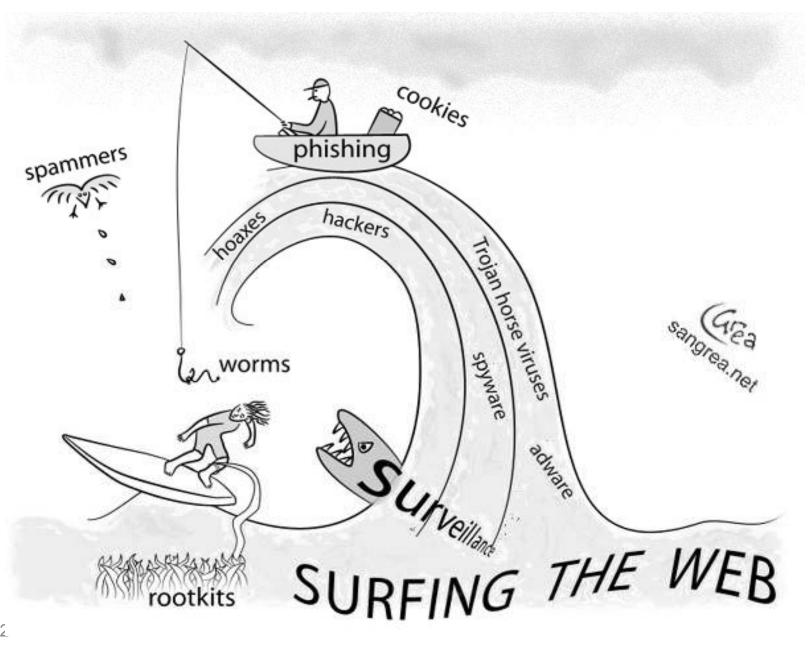
Are Google and Facebook Evil?

- Targeted advertising
- Cross-selling of users' data
- Personalized experience



Toothpaste For Dinner.com

Online Privacy



Some improvements on transparency

- An interview by Privacy International with Google on Government access to personal information, 2010
- Google transparency reports listing the requests received by Google from government entities for the disclosure of user data in six-month blocks.

Home

Traffic

Government Requests

▶ Country Detail

Map

Content Removal Requests Like other technology and communications companies, Google regularly receives requests from government agencies and courts around the world to remove content from our services and hand over user data. Our Government Requests tool discloses the number of requests we receive from each government in six-month reporting periods with certain limitations.

These numbers represent the requests we received from government entities for the disclosure of user data in six-month blocks. We continue to look for new ways to organize information and provide more detail. For example, starting with the July-December 2010 reporting period, we began to disclose the percentages of user data requests we comply with in whole or in part. And starting with the January-June 2011 reporting period, we began to disclose the number of users or accounts about which data was requested.



January to June 2011



User Data Requests

Raw Data

FAQ

Visible Changes

▲ Country	User Data Requests	Percentage of data requests fully or partially complied with	Users/Accounts Specified
Argentina	134	32%	188
Australia	361	73%	412
Belgium	90	67%	111
Brazil	703	87%	1,822
Canada	50	48%	75
Chile	118	42%	143
France	1,312	47%	1,552
Germany	1,065	66%	1,759
Hong Kong	123	42%	123
Hungary	81	2%	82
India	1,739	70%	2,439
Israel	60	60%	67
Italy	934	60%	1,263
Japan	75	87%	82
Mexico	48	42%	73
Netherlands	64	48%	213
Poland	266	11%	319
Portugal	161	50%	323
Russia	42	0%	47
Singapore	106	75%	126
South Korea	259	37%	530
Spain	460	63%	709
Switzerland	36	69%	42
Taiwan	155	81%	267
Turkey	73	0%	74
United Kingdom	1,279	63%	1,444
United States	5,950	93%	11,057

They are always watching ... what can we do?



Who cares? I have nothing to hide.

If you do care ...

- Use cash when you can.
- Do not give your phone number, social-security number or address, unless you absolutely have to.
- Do not fill in questionnaires or respond to telemarketers.
- Demand that credit and data-marketing firms produce all information they have on you, correct errors and remove you from marketing lists.
- Check your medical records often.
- Block caller ID on your phone, and keep your number unlisted.
- Never leave your mobile phone on, your movements can be traced.
- Do not user store credit or discount cards
- If you must use the Internet, encrypt your e-mail, reject all "cookies" and never give your real name when registering at websites
- Better still, use somebody else's computer

Privacy Protection Techniques

- Finding balances between privacy and multiple competing interests:
 - Privacy vs. other interests (e.g. quality of health care; movie recommendation)
 - Privacy vs. interests of other people, organization, or society as a whole (e.g. insurance companies, healthcare research; movie recommendation for others).

Security

- The quality or state of being secure: as a: freedom from danger; b: freedom from fear or anxiety (merrian-webster)
- National security
- Individual security
- Information security
 - Computer security
 - Data security

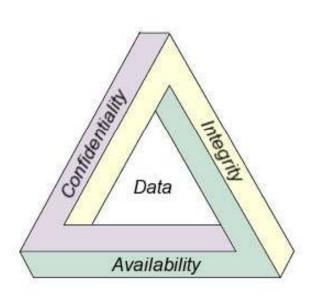
Security vs. Privacy

- Data surveillance
 - Surveillance cameras
 - Sensors
 - Online surveillance



Principles of Data Security – CIA Triad

- Confidentiality
 - Prevent the disclosure of information to unauthorized users
- Integrity
 - Prevent improper modification
- Availability
 - Make data available to legitimate users



Privacy vs. Confidentiality

Confidentiality

 Prevent disclosure of information to unauthorized users

Privacy

- Prevent disclosure of personal information to unauthorized users
- Control of how personal information is collected and used

Data Privacy and Security Measures

Access control

Restrict access to the (subset or view of) data to authorized users

Inference control

Restrict inference from accessible data to additional data

Flow control

 Prevent information flowing from authorized use to unauthorized use

Encryption

 Use cryptography to protect information from unauthorized disclosure while in transmit and in storage

Course topics

- Access control
- Inference control
- Secure multi-party computations
- Applications: healthcare, social networks

 Disciplines: databases, information security, data mining, statistics, cryptography

Access Control

- Identification and Authentication
- Authorization
- Access control policies
 - Discretionary access control
 - Mandatory access control
 - Role based access control
- Accountability and auditing

Security Measures

Access control

Restrict access to the (subset or view of) data to authorized users

Inference control

Restrict inference from accessible data to additional data

Flow control

Prevent information flowing from authorized use to unauthorized use

Encryption

 Use cryptography to protect information from unauthorized disclosure while in transmit and in storage

Inference Control

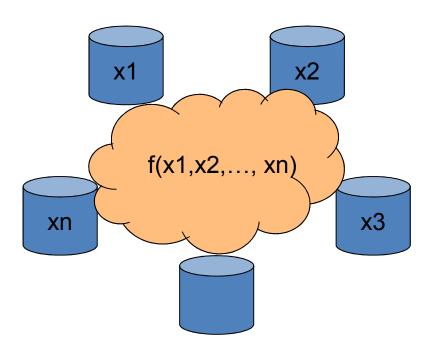
- Inference control: Prevent inference from deidentified, anonymized, or statistical information (accessible) to individual information (not accessible)
- Attack Incidents
 - Massachusetts Group Insurance Commission (GIC) medical encounter database
 - AOL search queries
 - Netflix prize

Inference Control

- Data anonymization
 - Data generalization
 - Data aggregation
 - Data perturbation
- Statistical database
 - Query restriction
 - Output perturbation
- Privacy preserving data mining
 - Data perturbation
 - Output perturbation

Secure Computations

- Multi-party secure computations
 - Cryptographic protocols
 - Absolute security/privacy vs. approximation



Today

- Meet everybody in class
- Course overview
- Course logistics
- Poll

Logistics

- Materials
 - Papers, online articles
- Prerequisite
 - Some database and statistics background
 - Programming skills
- Class webpage
 - Lecture notes
 - Link to readings
 - Project/assignments

http://www.mathcs.emory.edu/~cs573000

Workload

- ~2 programming assignments (individual)
- ~2 reading assignments
- ~1 paper presentation
- 1 open-ended course project (team of up to 2 students) with project presentation
 - Application and evaluation of existing algorithms to interesting data
 - Design of new algorithms to solve new problems
 - Survey of a class of algorithms
- 1 midterm
- No final exam

Late Policy

- Late assignment will be accepted within 3 days of the due date and penalized 10% per day
- 1 late assignment allowance, can be used to turn in a single late assignment within 3 days of the due date without penalty.

Grading

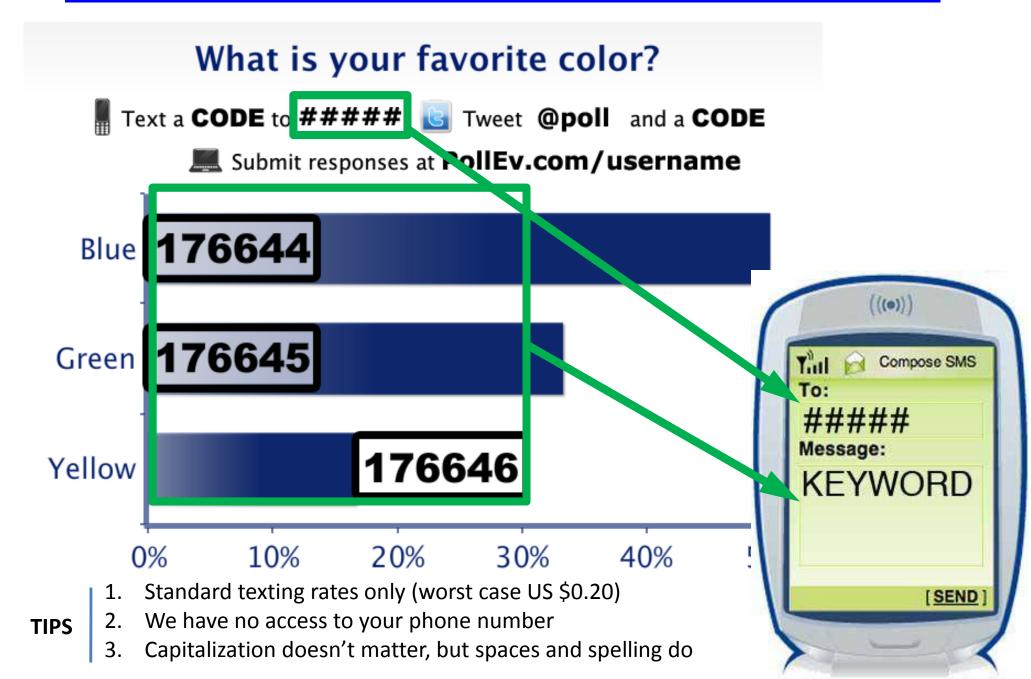
•	Assignments/	presentations	40%
---	--------------	---------------	-----

- Final project 30%
- Midterm 30%

And now ...

- Meet everybody in class
- Course overview
- Course logistics
- Poll

http://www.polleverywhere.com



Online recording

How concerned would you say you are with the following aspects of the Internet?

Companies recording your online habits and using the data to generate profit through advertising

Very concerned 44%

Somewhat concerned 37

Not very concerned 15

Not at all concerned 4

Not sure

Online tracking

Do you believe law enforcement should have to get a warrant to track where you go on the Internet, like they have to get one to wiretap phone conversations?

- Yes 79%
- No 12
- Not sure 9

Government for online privacy

Do you believe government regulators should play a larger role in protecting online consumer privacy?

- Yes 49%
- No 36
- Not sure 16

Online anonymity

- Statement A: "I think anonymity on the Internet has to go away. People behave a lot better when they have their real names down. ... I think people hide behind anonymity and they feel like they can say whatever they want behind closed doors."
- Statement B: "Many people believe that requiring real names will solve the problems of trolls and bad behavior, but they don't -- and that policy can have negative consequences in terms of suppressing dialogue about important topics."

1/25/2012 45

Online Anonymity

Which statement comes closest to your opinion?

Statement A: "I think anonymity on the Internet has to go away. People behave a lot better when they have their real names down. ... I think people hide behind anonymity and they feel like they can say whatever they want behind closed doors."

Statement B: "Many people believe that requiring real names will solve the problems of trolls and bad behavior, but they don't -- and that policy can have negative consequences in terms of suppressing dialogue about important topics."

•	Anonymity on the Internet has to go away	21%
•	Requiring real names suppresses dialogue	49%
•	Neither	19%
•	Not sure	12%

Online Privacy

Would you consider someone posting a picture of you in a swimsuit to be an invasion of your privacy?

- Only 35.6 percent of 18-24 year-old consider it an invasion of privacy
- 65.5 percent of other respondents