Privacy technologies grad seminar: Online discussion prompts

This document lists the online discussion prompts for privacy graduate seminar at Princeton, Fall 2012. See the link for the reading list.

0. Intro

Solove and Brin

1. In class we discussed the fact that Fifty Shades of Grey became a hit only because we can read privately now in a way we couldn't before digital books. What are some ways in which technology has improved privacy?

2. There is a widespread concern (example) that reader privacy is going away in the digital age. And yet as the Fifty Shades example shows, people are also benefiting from greatly improved reading privacy. What's going on? Can both be happening at the same time?

3. One way we're adapting to our information being increasingly made public, according to a common argument, is that the underlying social values that require privacy are themselves changing. For example, it used to be the case that you had to keep your sexual orientation private if you were gay, but that is less important now because society is more tolerant of different sexual orientations.

Now think about Solove's example of cancer, as well as other health conditions. How might society adapt so that it is less of a problem if our health conditions were made public? Do you think this kind of adaptation will actually happen?

4. Let's focus on the Kafka argument. List as many Government functions/agencies as you can that use data about citizens to make decisions that have a significant impact on their lives. In general, as these processes become more data-driven and automated, do you feel that the potential for Kafkaesque treatment has increased or decreased?

5. What do you imagine Brin would say in response to Solove's arguments?

6. Brin's essay was written in '96. In retrospect, what did Brin get right and wrong about the development of surveillance technologies?

7. What are some laws and/or current court cases that have a bearing on Brin's ideal of "watching the watchers?"

8. What are some technologies that can be used to achieve this kind of transparency?

9. Solove and Brin both focus on privacy w.r.t. Government. Why is this? What other types of privacy do people worry about?

1. Cryptography: power and limitations

Chaum

1. In the dining cryptographers example (first figure), why bother with a coin? Why can't the two friends confer between themselves and tell you if one of them paid?

2a. In Chaum's world, you go to the store, buy some clothes, and pay with your "card computer" via blind signatures without having to reveal your identity. Does the checkout terminal necessarily have to connect to your bank for the transaction to go through? Why or why not?

2b. The clothes don't fit so you go back to return them a few days later. How does the store verify that you are eligible to return them? (i.e, that you bought them at the same store within the return window.) Can they do this in a way that doesn't involve linking your pseudonym across the two transactions (purchase and return)? [Note that the more transactions your pseudonym gets linked to, the more anonymity you lose.]

2c. You go back again and try to return identical looking items that you bought at another store. How does the store prevent you from returning these? [Note that items don't have unique tags when they are returned because customers typically remove them.]

3a. You go to a bar and would like to be able to provide proof of age without revealing your identity. In Chaum's world this would be a digital credential. Would you rather this were a certificate of your birth date/year, or simply a credential that certifies that you're over 21?

3b. How would this credential be tied to your identity to prevent an analog of presenting a fake ID? There is more than one way. What are the advantages and disadvantages of each?

4. Chaum is focused on commerce. Let us talk about the other great realm that the Internet has revolutionized: communication and social networking. Under the crypto vision, Facebook wouldn't know all your secrets; at best it would be a "dumb" carrier of encrypted messages.

How would you encrypt your wall posts so that all your friends can read it, and no one else? How efficient is your solution in terms of bandwidth? If it is inefficient, can you improve it? What happens when you add a new friend — can they see you old posts? What if you have custom friend lists?

5. Most of what Chaum describes has an analog in terms of physical envelopes. Can you think of computations that crypto enables that are impossible in the physical world even using tricks

like envelopes and coin flipping?

6. Just as credentials are tied to identities in our world, so is reputation. There are many numerical reputation scores that follow us around and vary with time — credit score, karma on reddit, etc. So in a pseudonymous world we will need to find a way to reimplement reputations without identity.

6a. Let us consider a simple reputation system: whenever a certain event happens (e.g., paying a bill on time), an issuing authority (e.g., your bank) increases your reputation (by a variable amount depending on the type of event). Other parties (e.g., landlord) ask for proofs of reputation, i.e., that your reputation exceeds a certain threshold. How would this system implemented? Easier version: assuming that a credential system already exists, how can you repurpose it into a reputation system?

6b. What if your reputation also needs to be decreased sometimes? Do you think this is possible to implement?

7. Does this paper deal with cryptography for privacy, for security, or both? Give some examples of each (whether or not from this paper.)

8. How would Chaum's world be different from ours socially, politically, and economically? What would be the effect on currency? on trust, crime and punishment?

Chaum again and a bit of Levy

1. List as many examples as you can of the use of crypto in everyday life today.

2. What are some reasons why hasn't Chaum's vision come true? Is it more likely or less likely to come true now than it was in 1985?

3. Chaum devotes a single paragraph to the problem of abuse of a card computer by other individuals. Let's discuss this issue more thoroughly. Analyze the security vs. usability tradeoffs of the 6-digit PIN system. What are some other ways in which abuse could happen? What are some analogies in today's world from which we can draw hints about how things might turn out?

4. Chaum worries a lot about unconditional vs. computational security/untraceability. How important is this distinction? Do you know of examples where systems have been compromised due to the security guarantee being only computational and not unconditional (i.e., by factoring a large number, etc.)?

5. As Levy describes, the U.S. Government used to believe that cryptographic algorithms should be kept secret for maximum security. Do you agree with this policy? To what extent do

Government agencies still believe that? How do we know?

6. What impact did this policy have in hampering Chaum's vision?

Javascript cryptography and OtR

1. Each of the encryption tools PGP, OtR and cryptocat have a slow "key generation" process. Why is key generation slow?

2. Ptacek views security as all-or-nothing — as long as there is any viable attack, the system is completely worthless, perhaps even worse than useless. Is this a good principle? Argue for or against. Do you know of any examples of powerful entities attempting to or successfully compromising the web security of regular users?

3. The essay repeatedly emphasizes that secure delivery of Javascript to web browsers is a problem. Describe some concrete ways/attacks in which this security can be compromised.

4a. Imagine an encrypted chat application like cryptocat. It previously used to be purely web-based, and was only recently converted to a plugin-based app due to a chorus of criticism similar to Ptacek's. Assume that the web version can be built to deliver Javascript securely (by delivering the entire page over SSL, etc.) Is there still a problem with this security model?

4b. Can you think of any other applications where it actually makes sense to use Javascript cryptography, again assuming that it can be delivered securely?

5. What can Javascript do to generate random numbers more securely? Are there famous examples of security failures caused by insufficient randomness?

6. Think about email vs. instant messaging. In which application is authentication more important? Is there another form of communication in which it is even less important?

7. OtR claims to "append an identifier" to the first message to determine if the other party is using OtR. How come you don't see this identifier when you chat?

8. What happens if a user has multiple chat programs open, some of which do support OtR and some don't?

9. This is one of the key paragraphs of the OtR paper. Every sentence makes a nontrivial point about security. Explain each one. "During the initial Diffie-Hellman key exchange, we notify the user that we are about to start secure communication and display the fingerprint of the other party's public key. A more cautious user will verify the fingerprint out-of-band; for others, a man-in-the-middle attack is possible at this point. We record the public key value and make sure

the same key is used in future sessions. Thus, a successful impostor must be able to carry out an active attack during the first and every subsequent session; failure to do so will result in detection. This model of handling public keys is analogous to that used in SSH, and has proven an effective way to distribute public keys in the absence of a widely-deployed public key infrastructure. Of course, if a public key can be securely obtained by other means, it can be imported into the key store and used in the future"

PGP

1. List all the usability problems you encountered while installing, setting up and using PGP. What functions did you find difficult or impossible (or didn't even bother to try?)

2. How widespread is the use of PGP encryption these days? Which categories of users (if any) use it? Does this give you a hint about why the user interface remains so problematic for general users?

3. What are the different pieces of software/infrastructure in the PGP ecosystem? (There are more than you might at first think!)

4. How would you use PGP with Gmail? (There is more than one way.)

5. What entities do you need to trust in order to send PGP-encrypted mail securely? (Again, there are more than you might at first think, even after having answered Q3!)

2. Data collection and data mining, economics of personal data, behavioral economics of privacy

Price discrimination

1. Let's do a thought experiment on price discrimination. Think about a sci-fi world in which there were no real barriers to the practice. What are the sources of personal data that companies would use? (Especially data sources that aren't used today.) What sorts of inferences would they draw about people? And how would they price products using these insights? Finally, would you want to live in this world?

2. What are the main barriers to the above scenario in today's world?

3. Odlyzko mentions knowing the customer's identity as an important enabler of price discrimination (because it allows prohibiting resale, and resale would defeat the point of price discrimination). What are some current and potential technologies that businesses (could) use to determine the customer's identity?

4. How can the growth of online data collection feed into price discrimination in the offline world?

5. What are some ways of ‘covert’ price discrimination (used so that customers don’t perceive it as unfair)? Identify the examples in the paper, as well as others.

6. Analogous to railroad regulation, do you think consumers today will rebel against price discrimination and push to regulate it? Why or why not?

7. Consider the “sense of fairness” that is discussed extensively in the paper. Is there a rational economic model that partly explains it? Is it also in part hard-wired? What does this question have to do with Capuchin monkeys?

8. The NYT article mentions a list of 25 products that when analyzed together, allowed assigning each shopper a “pregnancy prediction” score. What is the process by which this list might have been derived?

Third-party tracking

1. One of the hotly contested issues in online tracking is whether tracking should be opt-out or opt-in, i.e., whether the default should be tracking or non-tracking. Does this actually matter, since both provide the same choice? Look up Coase’s theorem; what does it suggest as an answer to this question? How applicable is the theorem to this scenario?

2. Mayer and Mitchell are not happy with the studies on measuring the effectiveness of behaviorally targeted advertising. Describe the design of a better study that an ad network could carry out.

3. Experiment with Panopticlick.

3a. Try to minimize the identifiability of your usual browser or another browser. What’s the most anonymous you were able to get? With what settings?

3b. Are there tools you can download that are specifically intended to resist fingerprinting?

3c. Panopticlick uses things like the list of installed fonts for fingerprinting. However, these might change frequently. Describe an algorithm that allows for accurate fingerprinting when the information in question changes slightly but frequently.

3d. Are there applications of fingerprinting for fraud prevention?

4. Neither self-regulation in the U.S. nor Government regulation in the EU (e.g., the cookie law)

has worked particularly well. What are some reasons that these attempts have run into problems?

5. For “Do Not Track” to be meaningful, there has to be some way of detecting trackers that are not in compliance. What are some ways of doing so?

6. A frequent criticism of user surveys (of the kind cited in the paper showing opposition to tracking and behavioral advertising) is that user behavior is very different from these surveys — people give up their privacy with the slightest of incentives. Discuss this criticism.

7. Yao’s theorem assures us that if we can express behavioral advertising as a function of inputs belong to two parties (user information held by the user, ads held by the ad network) then we can do behavioral advertising with privacy. Describe such a function. Describe the architecture of a system based on this principle.

8. Speculate on what the state of online tracking might look like in 5 years.

Lemons and Social Networks

1. Lemons.

1a. There are several markets relevant to the context of online privacy (one of which is studied in the paper). List them. [*My goal here was to get students to realize that the market for personal data is different from the market for privacy tools. In retrospect this question was worded too ambiguously.*]

1b. Explain in detail how the lemons argument applies to privacy (the answer to 1a is relevant here). What are the differences compared to security? Can you think of any examples of products/services getting “priced out” due to good privacy practices?

2a. Do Bonneau and Preibusch have a normative stance on privacy? Justify your answer with quotes from the paper.

2b. What is the field of study in which this work is situated? When did this field get started and what are some of its founding assumptions?

3. Bad practices.

3a. At the time of the study almost half the sites requested a password to the users’ email account. What’s the current situation? How is it different and why?

3b. Think about it from the point of a social network. How would you respond to the criticisms in

Section 4.4?

4. Fundamentalism.

4a. List as many examples as you can of sites changing their privacy policies for the better due to complaints/actions of privacy fundamentalists.

4b. List as many examples as you can of efforts of privacy fundamentalists failing to have the desired effect.

4c. What are some organizations that are wholly or partly in the business of being privacy fundamentalists?

5. Evolution.

5a. How has the market evolved since the study (in terms of the number of competitors)?

5b. What explains the phenomenon of different social networks dominating in different countries?

6. Privacy controls.

6a. List as many different access control options as you can for limiting the visibility of profiles, posts etc. on social networks. (Feel free to list hypothetical ones.)

6b. Now try to go beyond the access control model.

Behavioral economics

1. What are some commercial practices or product features that are intended to exploit consumers' irrationality to entice them to give up their privacy? There are some examples in the reading; do list them for completeness, but also find others. This will require some work, for instance, playing around with websites, reading blog posts on things like 'gamification', etc. There are innumerable instances, so hopefully each of you will be able to find at least one new example.

2. What are some rational reasons to give up one's privacy? The answers can be context-specific.

3. In equation 4 in the immediate gratification paper, is β rational or irrational? What about δ? If it is rational, explain how it can be rational. If it is irrational, prove it.

4. List some ways in which time-inconsistent but sophisticated individuals protect themselves in

domains other than privacy (e.g, dieting). Can some of these techniques be the inspiration for privacy technologies?

5. The US Federal Trade Commission is a consumer protection agency tasked with investigating and preventing “unfair and deceptive” trade practices. If a company exploits consumer irrationality but does not make outright false statements, should they get investigated?

3. Anonymity and De-anonymization

Tor

These questions were supplementary to a more traditional paper discussion.

1. What are the different types of nodes in Tor? Explain them.

2. How many Tor relays are there, roughly?

3. Describe how an end-to-end correlation attack would work, verbally at a minimum, but if possible using a simple mathematical model for the number of nodes, amount of traffic, etc.

4. What are some of the various ways (dozens) in which the Tor browser blocks browser fingerprinting? You might need to look this up online.

5. a. If you try to log in to (say) Facebook using Tor, how does the site respond? Why? b. If a website chooses to block all Tor users, how can it do that?

6. What are the venues (journals/conferences/workshops) in which studies of Tor (and communications anonymity in general) are published?

Bitcoin

1. Landscape

a. Can digital cash system have all three of the following properties: decentralization, anonymity, double-spending prevention?

b. Give examples (including hypothetical) of systems that have each two, as well as (if you like) other digital cash systems with interesting technical properties.

2. Pseudonymity.

a. What is a bitcoin address, mathematically?

b. If this address doesn’t have the receiver’s identity, how does the sender know whom to talk to?

3. Blocks

a. Why do we need the block as a separate abstraction — why can’t each transaction have its own block?

b. How frequently are blocks generated on average? What is the distribution of inter-arrival times, i.e., time between two consecutive blocks?

c. Any node can start working on calculating the next block. So do nodes talk to each other to figure out who will do it? If so, how? If not, doesn’t it lead to wasted (duplicated) effort?

4. Theft

a. What are the two ways in which you can lose access to your bitcoins? Try to find real-world examples. [Edit: this question may have been unclear, I meant the two ways in which you can lose the ability to spend your bitcoins. One of them is losing your wallet, and the other is theft, which is what the next question is about.]

b. Suppose your private keys get “stolen,” i.e., copied by a thief. Now both of you have the same information. Then who has the money?

c. Let’s say you successfully get away with theft. What are the various steps you can take to hide your tracks?

d. Conversely, what are the various forensic techniques (based on future transactions that they have to make in order to spend the stolen bitcoins) that one can use to try to uncover the thief?

5. Economy

a. How many transactions have there been so far? How many public keys have participated in it? What is the current total value of Bitcoins?

b. What are some of the prominent bitcoin-based businesses/services and what are their functions?

More Bitcoin

1. Following up on answers to the theft question from last class, tracking of tainted Bitcoins and preventing them from being spent is extremely difficult, if not impossible. Explain why. (It might help to think through the exact algorithm you'd use to do this and figure out how it can be worked around.)

2. Describe a technical attack with which a powerful adversary (e.g., a Government) who controlled a lot of CPU power could destroy the Bitcoin system — e.g., cause it to lose most of its value and thus become useless.

3a. Check out Mt. Gox (and other exchanges if you're interested). Describe its technical architecture. (If it's not clear what that means, think of it as analogous to an abstract datatype. Describe its interface and a possible implementation.)

3b. Are Bitcoin exchanges susceptible to runs?

3c. Exchanges have proven to be extremely attractive targets for hacking because of the difficulty of catching Bitcoin thieves. What can operators of Bitcoin services do to minimize the damage from hacking, in addition to usual network security measures?

4. Describe the technical architecture of Bitcoin laundry, i.e., a mixing service.

Anonymity: ethics, law, economics

1. List as many good and bad uses that people have put anonymity tools to, especially Tor and Bitcoin, but also just anonymous publishing (e.g., Wikileaks). "Good" and "bad" are of course subject to your interpretation.

2. Discuss the ethics of online anonymity. Some questions to guide your thinking: a. How much anonymity "should" we have in an ideal world? b. How should anonymity be regulated by governments, if at all? c. What responsibilities do companies that control social media platforms have? d. What about developers of anonymity technologies?

3. How is Tor being used for circumvention of censorship?

4. What are the types of currency studied in economics? Does Bitcoin fall into one of these categories? Why or why not?

5. What are the different ways in which law enforcement can try to crack down on Silk Road, either the system itself or the participants?

6. What are some court decisions / legal precedents on compelling an operator of an online

service to unmask an anonymous poster who uses the service (say by providing the poster's IP address?)

Deanonymization algorithms

1. Human genome whirlwind tour: What is a base pair? What's a SNP? How many of each are there? How are the SNPs of a child dependent on the parents' SNPs? How is a population defined? What's an example of a population?

2. The goal in reading this paper is to be able to separate the genomics from the information theory. To this end, let's consider a simplified model:

Each human genome has exactly N=500k SNPs. SNPs are independent of each other. Each SNP is a single allele with a binary value, either 0 or 1.

Let's say there are M individuals whose SNPs are represented by the 2-D array y. Thus y[i][j] is a bit representing the value of SNP j for individual i. (Notation: if a is a vector, we will use a1 and a[1] interchangeably.)

a. DNA from the M individuals are present in a physical mixture in different proportions p1, p2... pM. Assume that genotype assays have enough probes to measure SNP frequencies accurately. Let m1, m2 … mN be the vector of frequencies measured. Express this vector in terms of the y's and p's. [This is intended to be a straightforward question, just to make sure you understand the notation and what an assay measures.]

b. Let pop[1..N] be the reference population allele frequencies. (One way to understand this is that if everyone's DNA in the population were present in a mixture in equal proportion, then pop[1..N] = m[1..N].)

Let x[1..N] be the SNP sequence of a target person whom you want to determine is present in the mixture or not. (You know x[1..N] because you already have the target's DNA. e.g., you're law enforcement and the target is a suspect and you got a warrant for DNA collection. The mixture was the DNA found at the crime scene.) Write pseudocode for determining if the target's DNA is in the mixture. Equation 1 in the paper is key.

c. Assume that every SNP has a probability 0.5 of being 0 or 1 in a given individual. Let's say x is indeed present in the mixture; x=y[k] for some k, so the proportion contributed by x is pk. What's the relationship between N and pk for the identification algorithm to be successful?

d. Assume that there's a cancer study with a case group and a control group, both of which have size M. The case group SNPs are y[1..M][1..N] and the control group SNPs are z[1..M][1..N]. As is typical in these studies, the researchers publish SNP-wise aggregates Y[1..N] and Z[1..N], where Y[j] = sum(y[i][j]), i.e., summed over the individuals, and similarly for Z.

3. Why are there so many authors in this paper? Do you know of papers with even more authors?

4. What were the consequences of the publication of this paper? How did the NIH react?

5. Ponder the essential equivalence between algorithms that perform forensics and algorithms that invade privacy.

6. Collectively write a survey of the deanonymization literature focusing on the algorithmic aspect — for each attack that's been published/demonstrated, describe the characteristics of the data, threat model (what the adversary knows and and what he is trying to find out), and the kind of algorithm used.

More deanonymization

1. Collectively write a survey of the deanonymization literature focusing on the algorithmic aspect — for each attack that's been published/demonstrated, describe the characteristics of the data, threat model (what the adversary knows and and what he is trying to find out), and the kind of algorithm used.

2. Error

a. Most reidentification algorithms have a nonzero probability of error. Does this make reidentification less of a problem? Why or why not? If targets of reidentification have deniability analogous to randomized response, does it nullify the privacy harm?

b. Cynthia Dwork has argued that the correctness of the claimed reidentification of the two individuals in the Netflix Prize database (with their IMDb accounts) does not matter — they are harmed by the resulting accusations in either case. Do you agree?

3. Expertise

A case based on Sweeney's reidentification research went to the courts. In Southern Illinoisan v. Illinois Department of Public Health, the court rejected the reidentification claim, arguing:

> "The court posited that it was not reasonable to believe that someone with less knowledge, education, and experience in this area would be as successful as Dr. Sweeney in using the information provided to arrive at the same results Dr. Sweeney reached."

> “Are there two people in the entire state of Illinois who could replicate Dr. Sweeney's results with the same limited data or are there two thousand? Are there zero or are there a million? These questions are significant because without some sense of the magnitude of the alleged threat of which the defendants complain, it is very difficult for this court to determine whether the data in question reasonably tends to lead to the identity of specific persons.”

Discuss the court’s reasoning.

b. If the expertise required for deanonymization of a particular type of dataset is rare, is it unethical to publish research on deanonymization of that dataset?

5. How have a. companies, b. policymakers reacted to deanonymization research and to calls to overhaul current practices (including essays like the Ohm paper)?

6. Self-revelation

a. Some have argued that in many cases like the Netflix study, reidentification typically requires the target to have self-revealed information publicly (e.g., on IMDb), that the numbers of people who do so are small, and that this behavior should not affect overall policy. “If I blog about a hospital visit, should my action render an entire public hospital admissions database (relied on by epidemiologists and health policy advocates) in violation of privacy law? Are the bounds of information flow really to be determined by the behavior of the most extroverted among us? This looks like a quagmire from which no reasonable normative position can emerge.” [Yakowitz, page 26.] Discuss this view.

b. On a related note, just how prevalent is the practice of online self-revelation of the sort that could serve as auxiliary information to deanonymize a database of medical records? (e.g., blogging about a hospital visit, reviewing a physician on Yelp.)

7. Ohm argues that banning deanonymization (even when the data is shared with a single entity and not with the public) will fail because the violation cannot be detected. Are there other laws that are similarly “impossible” to enforce because they ban computations or communications that people can theoretically carry out in privacy? How successful are these laws, given this impossibility?

8. What are some properties of a dataset that might make it easier or harder to reidentify?

9. The State Inpatient Databases are datasets containing (extremely sensitive) records of hospital visits, and are made available by most states to researchers. The AHRQ (Agency for Healthcare Research and Quality) which is the responsible federal agency that coordinates the data release seems quite cognizant of the reidentification risk and takes several steps to mitigate it. First, obtaining the data requires (physically) signing a data-use agreement which prohibits

reidentification. Second, completing an online Data Use Agreement Training Course is required. Third, there is a fee to obtain the data, although for some state databases including California, this is a token fee of $35. Finally, the analyst must describe the research project for which the data will be used.

4. Lightweight Privacy Technologies and New Approaches to Information Privacy

User-driven access control

These questions were supplementary to a more traditional discussion of the user-driven access control paper.

1. Pick a social network, or any other user-facing product or service where privacy is a concern, and analyze how it fares w.r.t. one or more of the five pitfalls. (As a simple example, Facebook has an icon for each post denoting public/friends-only/friends-of-friends/custom. This way, when you comment on a post you know who can see it. This reveals actual information flow, to the extent that users notice and understand these icons.) In particular, analyze the systems studied in the second paper — smartphone and desktop OSes, and browsers — *without* user-driven access control in terms of how they fare.

2. Check out this attack that was published recently. To defeat this type of attack, how can a browser create chrome elements or notification dialogs that can't be faked by a malicious page? How can a desktop OS create a dialog that can't be faked by a malicious application?

3. Let us try to explore the complexity and subtlety of people's mental access control rules, with location privacy as an example. Consider the question "with what granularity would you like to share your location, with whom, under what conditions and for what purposes?" List as many rules as you can that seem reasonable under this model.

Privacy through obscurity

1. Perhaps the most frequent counterargument to privacy-as-obscurity is that users should simply get used to privacy-as-secrecy, or privacy-as-access control (example). Discuss this view in two ways. a. In a normative sense, i.e., is this the right way to solve the problem? b. In a descriptive sense, i.e., will this ever happen, either by itself or with increased user education?

2. List as many examples as you can (excluding the ones in the paper) of obscurity-by-design in social media. Both existing examples and hypothetical design tweaks to existing products are welcome. Especially valuable are examples of lack of obscurity-by-design that led to user outrage at the service provider.

3. To what extent does obscurity-by-design play well with social media companies' business imperatives? (Either in general or w.r.t. specific design examples.)

4. Describe the standard critique of security through obscurity and discuss how relevant it is to the authors' proposal for obscurity-by-design.

5. The law paper seems to take an unabashedly normative stance on privacy. How can you tell? Is this stance necessary for the authors to achieve their aims?

6. The authors use the term identity differently from the way it is used in computer science (and in fact they use the term identification for the latter.) What is your understanding of what the authors mean by identity?

7. What are some other factors of obscurity (i.e., other than the four listed by the authors) you can think of in the context of: a) a legal test b) design of social media?

8. What are some of the weaknesses of the authors' proposal in the law paper? Relatedly, why has the public/private dichotomy been so popular in legal rulings?

9. You've read three law papers so far (Ohm, self-surveillance privacy, and this one). In each case, which kind or kinds of entities in the legal system do they hope to impact, and how?

Self-surveillance Privacy

1. The authors repeatedly state that they offer no normative account of privacy. What do they mean by that and why is it important to them?

2. The authors eschew describing the technical architecture of the system. Let us try to do that. Here are some concrete things to think about.

a. There are many parties involved in such a system: PDV, smartphone platform provider (if sensing is done on a phone instead of a dedicated device), creators of sensing apps, creators of other apps that use the data (3P-ASPs in the paper's terminology). How would all these interact with each other? What features would the APIs provide?

b. How does one deal with the various types of data that the system needs to handle? Does the PDV need to have logic for all these data domains/types or can it be left to applications?

c. What privacy technologies can be used and how? Would encryption be useful? What are some technologies that can be used for "Privacy Rights Management" suggested in the paper? How sound are these technologies?

d. Would competing PDGs be interoperable? If so, who will create the standard?

e. What are other technological hurdles that might arise, and how would you solve them?

3. What legal protections if any do we currently have for data stored in the cloud? What are companies' policies about this data?

4. In footnote 44 the authors talk about grey areas to their delineation of self-surveillance vs. third-party surveillance. Here is an example in practice: Zeo doesn't even allow you to see your own data without uploading it to Zeo's servers. Is this just an isolated example or are there serious problems with the ability to draw a clean line?

5. Look up regulatory capture. What are examples of regulatory capture in various domains? Could regulatory capture affect a Personal Data Guardian system?

6. Why is this paper so long and have so many footnotes?

Contextual integrity

Analyze various online services w.r.t. contextual integrity, with a focus on social networks. Pick a service, and perhaps a specific functionality; determine the relevant context, sender and recipient, subject and information type. Then think about what transmission principle might apply, whether the system in question is designed to conform to that principle, and how it can be improved. If you're not sure where to start, you might try to pick a specific product or feature launch that led to a public outcry — say Google Buzz — and try to explain the outcry through the lens of contextual integrity.

5. Purely technological approaches revisited

No Wiki discussion for this section.