Social Palimpsests - clouding the lens of the personal panopticon

Dave MURRAY-RUST [a], Max VAN KLEEK [b] Laura DRAGAN [b]

[a] *Centre for Intelligent Systems and Applications, Department of Informatics, University of Edinburgh*

[b] *Southampton*

Abstract. The use of personal data has incredible potential to benefit both society and individuals, through increased understanding of behaviour, communication and support for emerging forms of socialisation and connectedness. However, there are risks associated with disclosing personal information, and present systems show a systematic asymmetry between the subjects of the data and those who control and manage the way that data is propagated and used. In this chapter, we explore a set of techniques for ameliorating the tension between the desire for the benefits of sharing and a distrust of those with whom we share our data.

Keywords. Obfuscation; Data politics; Personal Data Stores; Social Machines; Pants-on-fire;

Introduction

0.1. The rise of personal data and services reliant on it

“Refusal is not a practical option, as data collection is an inherent condition of many essential societal transactions” [2]

0.2. Data, Capta, Fiat Data, Exhaust Data - and the impossibility of limiting data diffusion

Increasingly, in order to utilise services, we must provide our data to third parties. This ranges from mobile phone numbers being required for Yahoo accounts, to location data being shared with Foursquare or Grindr, to the NHS adding personal health information to centralised databases. We call this fiat data - an organisation uses its position to demand (by fiat) the disclosure of certain information in return for use of its services. In some cases, this is a necessary requirement for the service to be worthwhile, but in many cases it represents an attempt by the organisation to create a monetizable product from its users.

Sharing data, by definition, is the entrusting of other parties with information; this necessarily involves relinquishing control over how it is subsequently handled and disseminated. However, data is persistent, while people and contexts change.

A government may decide to share previously confidential data, as in the case of the recent care.data fiasco in the UK; a company can be bought and its assets acquiredthe purchase of Moves by Facebook raised issues around the terms and conditions of data handling companies; and even without malice accidents can expose vast swathes of personal data, or court proceedings may force private communications to become public - the Enron emails still represent the largest publicly available corpus of private emails.

0.3. Multiple Identity

A natural part of online life is the ability to tailor the persona we present to different communities and contexts. An individual may want to disclose certain things to their professional colleagues, while presenting differently to friends and family or non-mainstream friend groups. Sharing certain personal data is a barrier to this, as its basis in physical fact provides multiple opportunities for joining up otherwise separate databases. Most, if not all, social interactions involve both strategic omissions and various kinds of lies and non-truths to manage the myriad conflicting social demands placed upon us. The lie maintenance required to avoid discovery may be trivial ("sorry, Im hungry, have to go!") but may become significantly more complicated as lies extend over time, and become woven into the social fabric.

0.4. Why doing it socially is difficult

0.5. The case for lying and the importance of anonymity

0.6. Personal Data Stores - allies on the intimacy battleground

Personal data stores (PDS) represent a partial solution to issue of presentation: having trusted, user controlled repositories for data enables a more user-centric approach to management of capta—those data which we choose to take and preserve. Bridges can then be built between personal data stores and the rest of the world in order to support the connected, networked interactions which users now expect. If these bridges simply share the data, even in a controlled manner, nothing has been gained; hence the bridges become conduits for manipulating truth and constructing falsehoods. As personal data stores accumulate more real-time contextual data about the individual, as well as about the individuals social connections, PDSes can provide support for the often stressful and mentally burdensome task of lie maintenance, for example: i) identifying when a person's real activities or whereabouts contradict a lie, and might be discovered; ii) identifying indirect social channels that could expose a lie (e.g. through friends of friends); iii) suggesting appropriate lies to use which are least likely to be detected; iv) suggesting individuals to lie to to support lie maintenance (e.g. friends of the person being lied to)

0.7. Why verification and provenance are better than sharing

Sharing is a crude mechanism. Once data has been shared, the originator can no longer exert control over it, and must rely on the behaviour of the recipient, which as noted may fail to meet user expectations. Validation, however is a more subtle tool: if a users personal dataset can be made sufficiently questionable as to be useless on its own, then locus of control shifts to the user choosing to validate parts of the dataset, which can be performed in a more nuanced, contextualised manner. If a user is the final arbiter of trust, they can decide to i) sign parts of their record, so that it is verified public fact; ii) co-sign it with another entity, so either can verify it but not anyone else; iii) verify it through an anonymous channel, so that the entity to whom they provide verification cannot propagate the claim further. This verification can be carried out entirely separately from the datastore itself, allowing for the presentation of different datasets as valid in different contexts, as well as unorthodox methods such as using the Bitcoin blockchain to notarise datasets, so that they can be verified in the future without revealing them as true at the time.

1. Review of current approaches and tools

Types of disinformation [1]:

redaction is hiding some or all of the information in a message

airbrushing is changing some of the information. *local crowd blending* means change it to a nearby message likely to be plausible. *global crowd blending* means change it to a message in a dense part of the space.

curveball add extra distracting information, push message into low density space

2. A selection of PDS enabled obfuscation strategies

Alexander's taxonomy [1] discusses several types of disinformation which relate to modifying single messages. In contrast, due to the pervasiveness of modern communications, we are concerned with modifying message *streams*, where a trace of multiple values must be considered. Additionally, there is the possibility of interaction with others, whether it is collusion to strengthen obfuscatory practices, or the addition—purposeful or otherwise—of information which exposes the obfuscation.

It is problematic to consider the obfuscatory tactics here without a sense of the scenario in which they are to be deployed. Our scenario in this paper is:

> The user wishes to make use of services which expect location information; the location information provided is shared publicly and is almost certainly stored indefinitely. At times, the user may want to draw on location based information—such as restaurant recommendations or directions—and there may be times when they with to verify that they were at a particular location.

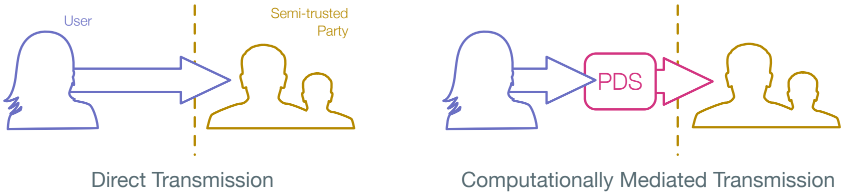


Figure 1. Models of interaction with semitrusted services. a) Direct transmission of information; b) computationally mediated transmission, where a personal data store is enlisted to aid in obfuscatory processes.

The service is hence *semi-trusted*: there are some benefits which the user wishes to accrue, but there are aspects of the service which makes the user unwilling to entrust their complete life history to it.

The standard model of interaction (Figure 1a) involves the user submitting their data directly to the service; for our obfuscatory techniques, we would like to enlist computational support (Figure 1b). This typified, but not limited to mediation from a PDS which acts on behalf of the user to modify the data which they provide.

2.1. Single Player Strategies

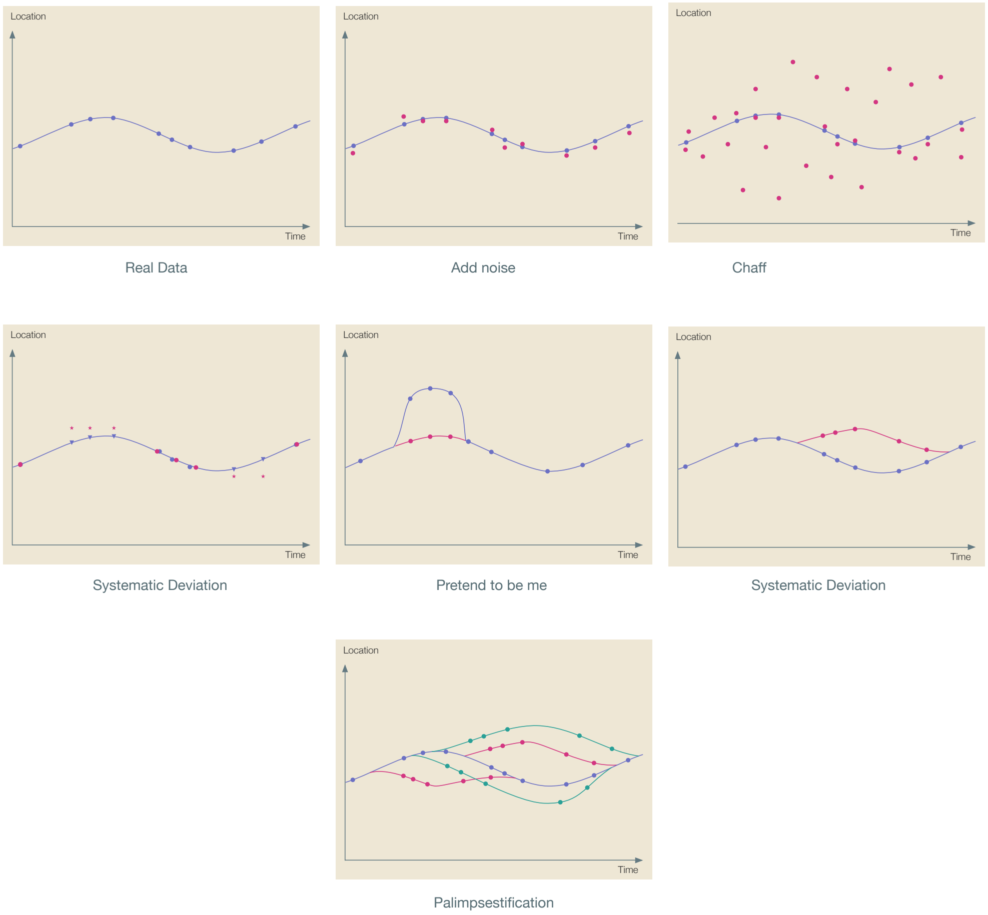
Figure 2 lists a collection of possible obfuscation strategies. In all cases, a fictitious one-dimensional "location" measurement is plotted against time, to give a sense of how an individual's position in space changes. Figure **??**a is the true baseline, with a curve indicating the continuous true position, and the dots representing reports of this position to the location-aware service. For each strategy we discuss: *i*) what kind of alteration of baseline data is performed; *ii*) what the motivation and possible use cases are; *iii*) how some form of computational support aids in the deception; *iv*) some of the systems which do this currently.

2.1.1. Noise injection

The most computationally simple form of obfuscation is the addition of noise to the reports which are sent to the semi-trusted party. Here, the points which are submitted deviate from the true values in a random manner. This allows the user to conceal their exact location, while giving a broad indication of where they are. **TODO:** *Why?* Avoid paparazzi/stalkers; conceal preferences - know you're in the high street, but can't tell if you're buying texbooks in Waterstones or beer in the White Horse

2.1.2. Chaff

- Add in lots of extra points, with little relation to the real values
- Computationally easy
- Can include the real points
- Disturbs services which expect a single consistent trace



Real Data

Add noise

Chaff

Systematic Deviation

Pretend to be me

Systematic Deviation

Palimpsestification

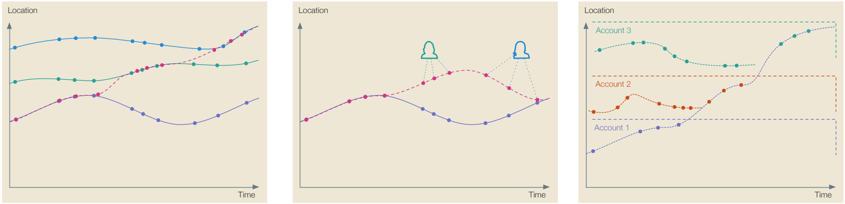
Figure 2. Obfuscation strategies for the lone agent

2.1.3. Systematic Deviation

- Modify specific points
- e.g. hid that we're going to ballet lessons, and pretend we're at the boxing gym instead
- Computational support: Can use services to find most useful nearby places to pretend to be at; allows for automatic, thematic deviations

2.1.4. Pretend to be me

- Generate "user-plausible" data, while the user deviates from their norm
- For instance, pretend to be at home while actually being on holiday
- CompSupp: have a model of normal behaviour which can be used to generate fake data



Artificial co-location

Supporting evidence

Account Sharing

Figure 3. Multiplayer obfuscaction strategies: i) artificial co-location; ii) supporting information; iii) account sharing

2.1.5. Virtual Holidays

- want to pretend that we're deviation from normal when we're not
- i.e. pretend we've gone to a conference when we haven't
- CompSupp: create a theme and help manage the deception

2.2. Multiplayer Obfuscation

2.2.1. Artificial co-location

- We can pretend that we're in the same place as our friends
- CompSupp: need to coordinate lies with friends

2.2.2. Supporting Evidence

- We ask our friends for supporting evidence of our lie; could be like co-location, could be broader

2.2.3. Account Sharing

- A group of people share accounts, with some way to decide which is used, e.g. accounts covering geographic areas
- Like swapping supermarket discount cards
- CompSupp: which account, when? discovery and sharing etc.

2.3. Verification and provenance mechanisms

Previously, we said verification is better. Why is this?

- Lets assume that we've made our public data completely unreliable so noone can use it.
- If someone wants to engage with it, they have to talk to us
- we can claim a subset of the data - just what they say they need for the purpose at hand
- We then have a choice of how to respond:
- No signing: if we simply send them a message saying "these points belong to me", then they can use the data, but would not be able to convince third parties of its validity.

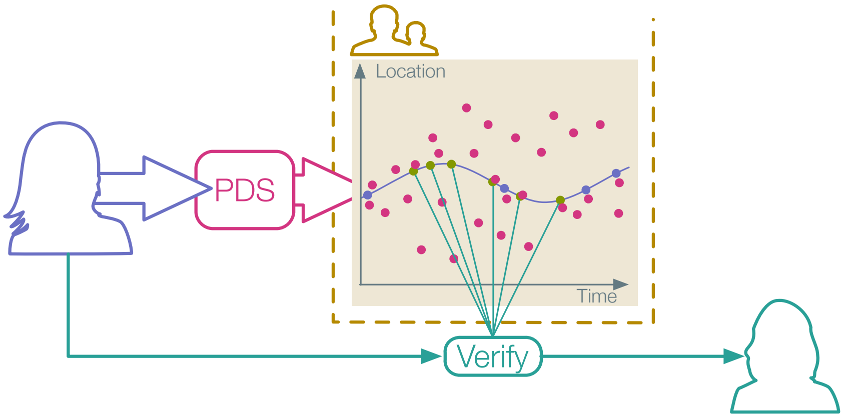


Figure 4. Example verification scenario. The user provides a set of real data, plus *chaff* to a location aware service. A third party then requests verification of some of the points, which the user provides

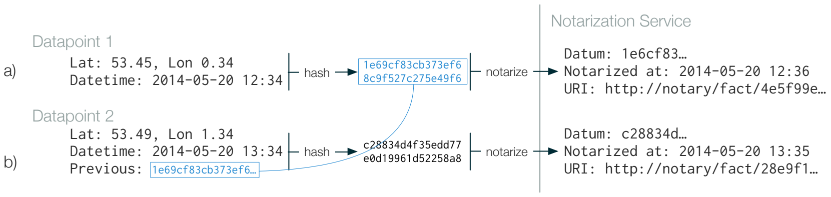


Figure 5. Notarization of personal data. a) Datapoints and times are hashed, and the values sent to a notary service, which provides a URL to verify that i) the given data was supplied and ii) when it was supplied. Hashes are used so that the data is not publicly shared. b) If the hash of the previous submission is included, then sequences of consecutive points can be verified.

- sign with our private key; now, anyone can check that we have claimed that set of datapoints
- sign with our private key and their public key: they can't prove the data is ours without revealing their identity.

2.3.1. Notarization

Third party notarization services can be employed. They take in some datum, and provide a hash/URL which can be visited to verify that that piece of data was provided at a certain time. For example, if someone wants to make a prediction for the outcome of a football match, they could notarize that before the match, and then subsequently prove that they had made the prediction beforehand. It is generall not possible to prove that they only made a single prediction, however, so this technique is most suitable when the range of possible things to notarize is so large as to make notarizing the entire space infeasible.

With regard to personal data, we can notarize our true data stream as we produce it. This means that we can prove that we had considered those points at the time, and if we say we were in a particular place, there is a high chance we were—however, it does not work in the complementary situtation as producing a notarized point does not prove that we were not anywhere else.

Notarization does not necessitate revealing the data itself. For instance, when submitting a location, a representation of the time and place could be hashed, and this hash notarized (Figure 5a). Additionally, points can be notarized in sequence, so that we can demonstrate contiguous sub-sequences of points as having been provided previously; by hashing the current location with the previous location, we can link the points together, to build up confidence in the notarized results (Figure 5b).

3. Conclusion

- Translation to things that aren't location data; generalisability; can't add chaff to our bank accounts (or can we?)
- Viability - how do services react when we fill them full of noise? Plausible versions of these techniques

References

[1] J. M. Alexander and J. M. Smith. Disinformation : A Taxonomy. Technical report, University of Pennsylvania Department of Computer & Information Science, 2010.

[2] F. Brunton and H. Nissenbaum. Vernacular resistance to data collection and analysis: A political theory of obfuscation. *First Monday*, 16(5):1–16, 2011.