# 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 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

Tools for preserving privacy have been decried repeatedly as methods for enabling bullies, criminals, miscreants, and even terrorists [] to more easily conduct their misdeeds against society. The "I have nothing to hide; therefore I should not care" argument [] used often in defense of ignoring privacy in the early days of targeted advertising and the Web still re-surfaces today. But as theoretical arguments of what *may* become have been replaced by tangible examples of what has actually has, this, among other arguments have become thoroughly dismantled. The sophisticated methods being used to track and pinpoint details of individuals has not been met with a corresponding increase in counter-tracking and counter-surveillance tools; as a result, people are being subject to an unprecedented barrage of targeted attempts to behaviourally manipulate them - often in ways too subtle for them to perceive (e.g. [**?**]).

Thus, this paper takes the position that privacy protection not merely as a concern for a minority with peculiar sensititivities towards being seen or exposed, but instead, as a set of capabilities necessary for preserving individual autonomy and well-being, in an economy where personal data has become the most precious digital commodity.

In this paper, we discuss and analyse methods of privacy protection that advance beyond the current state of anonymisation tools that merely obscure the

tracks of individuals towards mechanisms towards those employ methods borrowed from information warfare [], in order to allow individuals to regain autonomy from systems of behavioural control. We first discuss X and Y, followed by a a survey of lying and falsification in context-aware systems, current anonymisation ad privacy tools. This is followed by a complete overview of strategies

*The rise of personal data and services reliant on it, and relation so surveillance*

As we pass through the digitally augmented world that we collectively inhabit, the set of actions with the potential to produce data grows year on year. Portions of this outpouring are kept and stored as *capta*, from *capere*: to keep [**?**]. From using an access card to unlock a door at the workplace, right down to tracking individual footfalls, pervasive digital systems illuminate and annotate our physical activity. Accreting around this body of physical observation is an expanding sphere of mental observation and analysis. This can take the form of active practises around recording mental states, such as journalling, but it can also include computational inference, where frequency of posting on social networks becomes an adjunct metric for connectedness, and search terms indicators of intent. As such, the modes of collection of this information can range from explicit, user initiated submission of data, through consensual background recording to invisible, asymmetric electronic surveillance.

The pervasiveness of computationally mediated interaction means that for much of this data "refusal is not a practical option, as data collection is an inherent condition of many essential societal transactions"[**?**]. This leads us to introduce the term *fiat data*—when an organisation uses its position to demand (by fiat) the disclosure of certain information in return for use of its services. The term comes from a loose analogy with *fiat money*: currencies whose value is derived from the mandate of a government. **TODO:** *this is gibberish, needs rewriting!* Fiat money is, by definition *inconvertible*, and *intrisically useless*, and hence must derive its value from external forces, typically government decree, status as legal tender or by being the required currency of taxation. While the conditions of inconvertibility and uselessness are not necessarily applicable to personal data, the mechanism of central authority decree is commonplace; if one wants to interact socially on Facebook, one must pay the tax in personal data which they demand. This can be contrasted with modern cryptocurrencies such as Bitcoin; here, there is no central authority to backstop value and demand taxation. Rather, the participants in the economy collectively decide what the units of currency are worth; this would be a personal data economy in which participants decided what and how they wanted to share.

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.

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. One reaction would be be to reject the services which require data; to *opt-out*, but as Brunton and Nissenbaum put it: [**?**]

Gamification uses surveillance to effect behaviour change: "1) that gamification is a form of surveillance; and 2) this surveillance is pleasurable" from Gaming the Quantified Self [**?**], which leads to Albrechtslunds' "participatory surveillance"[**?**]:

> Online social networking can also be empowering for the user, as the monitoring and registration facilitates new ways of constructing identity, meeting friends and colleagues as well as socializing with strangers. This changes the role of the user from passive to active, since surveillance in this context offers opportunities to take action, seek information and communicate.

*Sharing is a loss of control*

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.

Brad Templeton: "Beware of time travelling robots from the future" or "Would you have liked to be gay 40 years ago in a monitored society? Or an enemy of J. Edgar Hoover with modern tools in his hands?" [**?**] Future tools will enable greater extraction and analysis of data after the fact;

David Lyon: leaky containers "data move freely between different sectors of society with the result that information from discrete contexts, e.g., private life, work life and shopping, are being mixed rather than contained separately."

*The case for lying and the importance of anonymity*

White lies can aid in social network growth: [**?**]

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.

Butler Lies: [**?**]

Translucence: "What we say and do with another person depends on who, and how many, are watching." - [**?**]

Contrast with Transparent society [**?**]; power imbalance between parties.

*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. **TODO:** *Content creation on youtube*

**TODO:** *Ben Dalton's thesis about Persona's throughout the agents*

*Why doing it socially is difficult*

Sharing certain personal data is a barrier to anonymity and obfuscation; data which is rooted in physical fact provides multiple opportunities for joining up otherwise separate databases. 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. The ability to compare multiple accounts of history—especially once the time travelling robots are involved—means that dissonance within the social fabric is more obvious than weaknesses in a single thread.

### *0.1. 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)

### *0.2. Why verification and provenance are better than sharing*

"Any model of privacy that focuses on the control of information will fail." "I call this practice social steganography. Only those who are in the know have the necessary information to look for and interpret the information provided." dana boyd [**?**].

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

**TODO:** *Max to do some writing*

- The revolution has started!
- TOR, anonymous remailers, burner phones, gotta change up, yo!
- Obfuscation symposium: `http://obfuscationsymposium.org/`
- Surveillance and Society: `http://library.queensu.ca/ojs/index.php/surveillance-and-society`

Theory of obfuscation: Types of disinformation [**?**]:

**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

Some existing stuff and the things we can link it to later

- TrackMeNot generates plausible google searches (Chaff)
- FaceCloak? Encrypts facebook data
- CacheCloak - sends a range of plausible future paths to location based services (Palimpsestification)
- Shopping card loyalty swaps (Account Sharing)
- DuckDuckGo - mixing up user searchers (Account Sharing, no cleverness)
- CVDazzle

## 2. A selection of obfuscation strategies

**TODO:** *Laura to write descriptions and try redoing with maps* Alexander's taxonomy [**?**] 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:

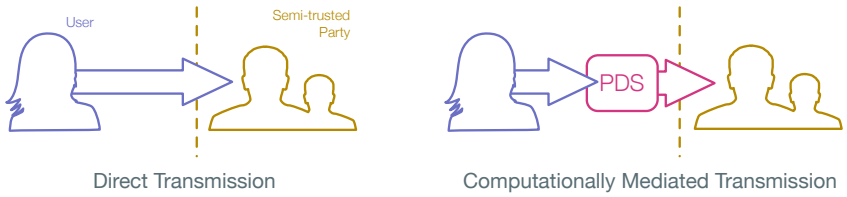


**Figure 1.** Models of interaction with semitrusted services. a) Direct transmission of information; b) computationally mediated transmission, where a personal data store is enlisted to aid in obfuscatory processes.

> The 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.

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. Strategies for the lone obfuscator*

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 2a 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*) how the strategy can be applied to other data *v*) some of the systems which do this currently.

#### *2.1.1. Chaff*

World War II fighter planes would emit clouds of radar reflective sheets—*chaff*—which created multiple traces the screens of radar operators, and hence disguised the true position of the aircraft. In a similar manner, we can add in multiple location datapoints alongside the real ones. This is the one of the few methods where the complete, accurate datastream is stored. Hence the user can still access any benefits which rely on accurate information. However, adding a multitude of randomised points to a service which expects a single contiguous trace is both easily detectable, and may break functionality—a run tracking application would be likely to give unreliable distance estimations in the presence of chaff.

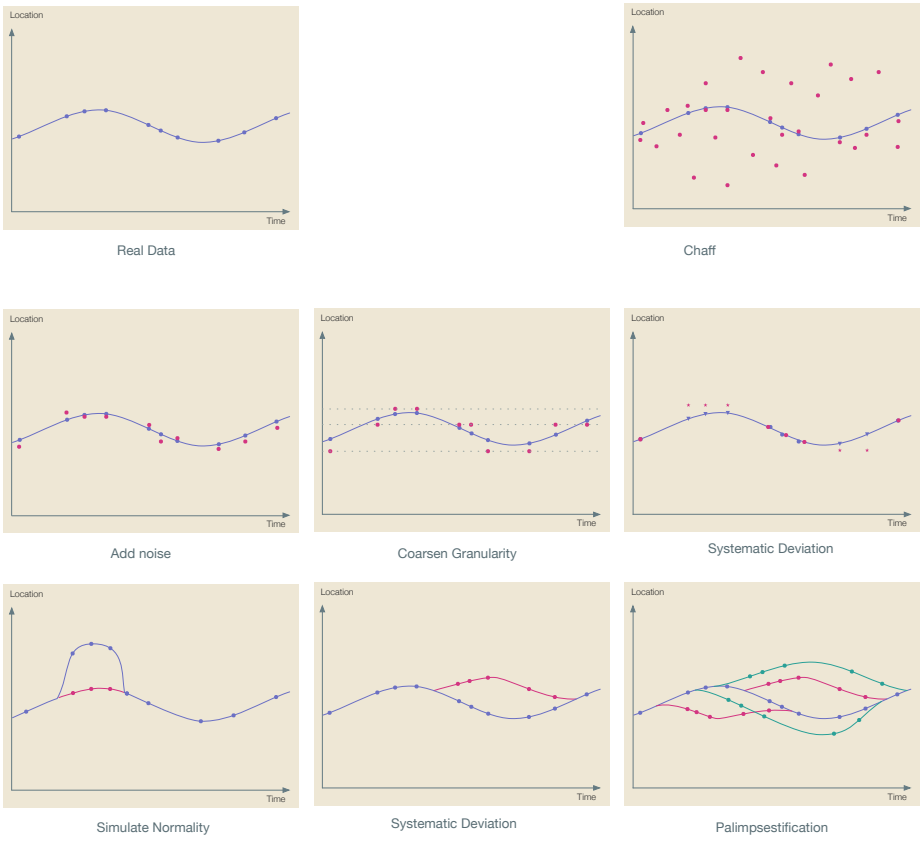


**Figure 2.** Obfuscation strategies for the lone agent



### *2.1.2. 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. Depending on the level of noise, this can allow the use of location based services without revealing much about actual behaviour. For example, it might reveal your location on the high street so you can arrange to meet friends, without revealing which shops you were visiting. This is compatible with services which expect coherent location data, and may be indistinguishable from the inaccuracies of the location sensors. One downside is that the "true" location traces are not present in the record of the service. **TODO:** *examples?*

### *2.1.3. Coarsened Granularity (or Quantisation)*

Rather than adding noise to the data being sent, it can instead be quantised to a coarser granularity, akin to blurring, or zooming out on a map. Again, this

is a technique which may help to derive useful information from the service, without revealing more than is necessary: using a service to find friends in the same city should only require city level information to be shared. An example of this can be found in Android's permission system, which has separate controls for `ACCESS_COARSE_LOCATION` versus `ACCESS_FINE_LOCATION`; similarly, posted letters may be signed with a city rather than a street address.

#### *2.1.4. Systematic Deviation*

In some cases, it may be possible to introduce systematic deviations into the digital record. In order to this, the user needs to be able to define which points to alter, and what to replace them with. One possibility would be thematic replacement—"hide the times I went to the pub by saying I was at a cafe". It is likely that this will require some form of computational support to i) identify targets for replacement as they occur and ii) find suitable replacements. Using this technique, some, but not all of the true data is stored; however derived information—such as beverage preferences in the example above—can be wildly and purposefully distorted. The nature and fact of the distortions may be hard to uncover, as no simulatneous traces or strange movement patterns are produced. Depending on the domain, subtle alterations may have large effects. **TODO:** *examples?*

#### *2.1.5. Pretend to be me*

With increasing computational support, it becomes possible to create a model of the user which outputs plausible "normal" data. This can then be used to replace periods of abnormal behaviour, or even replace normal behaviour with statistically similar but untrue data. An early example is when neighbours (or automatic switches) are employed to turn lights on and off in a home which has been vacated for the holiday, disguising the true anomalous data of a dark, empty house with the appearance of normal occupation. Similarly, one might avoid making Facebook posts which indicate an absence, to avoid burglary. This kind of deception can be difficult to achieve; however computational systems are emerging which can aid users, for example Beyer's digital alibi system [**?**].

#### *2.1.6. Coherent Deviation*

As the converse of simulating normality, the user may wish to pretend to be somewhere where they are not **TODO:** *more motivation for Alibot!* . This is similar to creating systematic deviations, but on a grander scale; the user would like to create a narrative for the deviation, and then have suitable data points constructed. For example, the use might pretend to be on holiday, or at a conference, and would like location traces which match that narrative to be created, such as going to the convention centre in the day, and returning to a hotel at night. This requires a computational model of user behaviour which can be applied to new locations—a non trivial task. However there is the potential to create obfuscated data which is difficult to distinguish from standard behaviour. **TODO:** *examples*

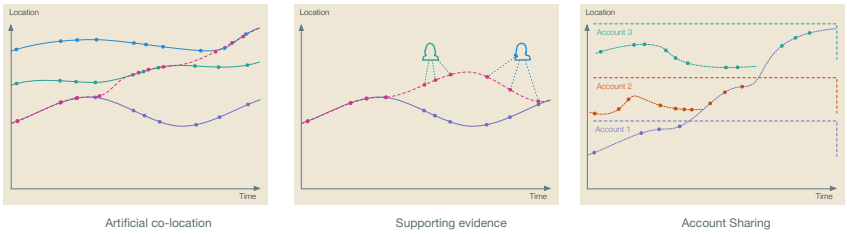


**Figure 3.** Multiplayer obfuscaction strategies: i) artificial co-location; ii) supporting information; iii) account sharing

#### *2.1.7. Palimpsestification*

Taking the idea of coherent deviations a stage further, and combining with the idea of *chaff*, the user could create multiple overlapping traces; each trace would be locally coherent and plausible, but someone inspecting the data would have no way to know which is the real one. This is similar to the strategy of CacheCloak **TODO:** *reference* , which continuously generates sheaves of probable future behaviour and searches location based services relevant to each path. The computational support required is similar to the coherent deviation example—to be able to run a model of the user's behaviour in novel locations—although more coordination might be required between the stories. The tradeoff is that while the true location data can be entered along with the generated points, the deception is obvious, and location based services may become upset at the multiple paths.

### *2.2. Collaborative Obfuscation*

Including others in the obfuscation challenge opens up a range of new strategies, where collusion can aid in the creation of otherwise unachievable datastreams, or increase the veracity of artificially created data. Generally the possibilities in computational systems are analagous to pre-computational possibilities; computational support tends to be in the form of coordination to find collaborators or and check coherence of data points.

#### *2.2.1. Artificial co-location*

One way to obtain a realistic but untrue location trace is to re-present the trace of a collaborator. This can look like relatively natural behaviour; two people meeting up to carry out joint activities or socialisation. Computational support here can involve finding accomplices to "co-locate" with—people who are willing to share their location, and are behaving in ways which match the desired story—as well as the technical business of transferring location devices between accounts.

#### *2.2.2. Supporting Evidence*

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

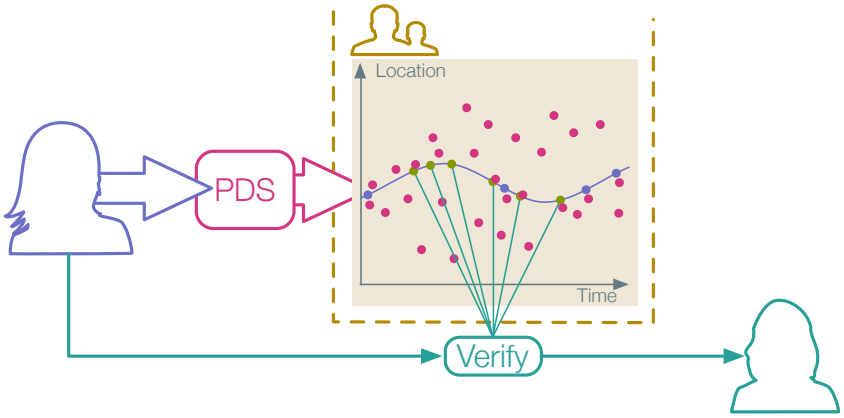


**Figure 4.** Example verification scenario. The user provides a set of real data, plus *chaff* to a location aware service. A third party then requests verification of some of the points, which the user provides

#### *2.2.3. Account Sharing*

In a similar manner to the swapping of loyalty cards discussed in [**?**], users of services can share accounts.

- 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.

## 3. Operationalisation - managing deception and its side effects

### *3.1. Going beyond location*

### *3.2. PDSs to support obfusction*

- create continuity
- act on your behalf

### *3.3. 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:

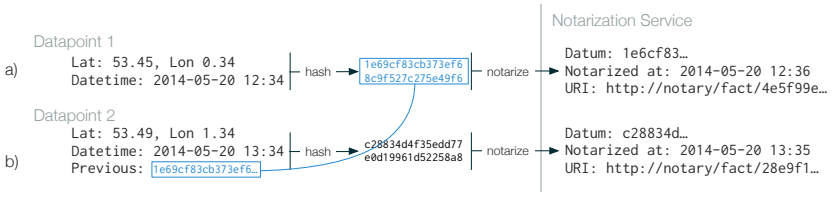


**Figure 5.** Notarization of personal data. a) Datapoints and times are hashed, and the values sent to a notary service, which provides a URL to verify that i) the given data was supplied and ii) when it was supplied. Hashes are used so that the data is not publicly shared. b) If the hash of the previous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

- 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.
- 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.

### *3.3.1. Notarization*

Third party digital notarization services can be employed[1]. These services take in some document or datum, and provide a certificate which can be used 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 generally 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).

[1] e.g. `http://virtual-notary.org/`, a free service hosted at Cornell University

## 4. 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
- Ethics - is this OK?
- Obfuscation evolves in lockstep with systems to see through it; future people will be better at spotting constructed points.
- In the short term services will start to become more suspicious about the data that goes into them; start rejecting points which represent causality violations.