Technical Sciences
Eemsgolaan 3
9727 DW Groningen
P.O. Box 1416
9701 BK Groningen
The Netherlands

www.tno.nl

T +31 88 866 70 00
F +31 88 866 77 57
infodesk@tno.nl

TNO report (35627)

Revocable Privacy 2011 – use cases

Date	5 Januari 2012
Authors	Wouter Lueks
	Maarten H. Everts
	Jaap-Henk Hoepman
Reviewer	Johanneke Siljee
Number of pages	22
Number of appendices	0
Project name	Revocable Privacy
Project number	035.33565

Contents

1 Introduction

This report is part of the Sentinels project for Revocable Privacy (project number 10532). Sentinels is partially funded by STW, NWO and the ministry of Economic affairs. The goal of this report is to give an overview of use cases in which revocable privacy can be applied. This overview is not exhaustive in the sense that it does not contain some well-known existing cases, nor does it even approach the actual number of use cases that are applicable.

The idea for revocable privacy stems from the ongoing conflict between security on one hand, and privacy on the other. It is widely believed that these two are mutually exclusive. The goal of revocable privacy is to show that this is not necessarily the case, and to design systems that offer both privacy and security.

In the original project proposal revocable privacy is defined as follows:

> *Definition*. A system implements revocable privacy if the *architecture* of the system guarantees that personal data are revealed only if a predefined rule has been violated.

In other words, users of the system remain anonymous unless they violate a predefined rule. The definition explicitly requires that the *architecture* enforces this rule. The naive solution to the problem is to simply store all necessary data and then later check if any rules have been violated. Procedural measures to restrict access to these data and hence protect privacy are, however, generally not good enough [1]. Hence, the naive solution is undesirable. In this document we will explore cases that could benefit from such an architecture.

Historically, one of the first examples of a system that implements revocable privacy is the electronic cash system proposed by David Chaum [2]. In this system, cash is represented by electronic coins. Since they are electronic, it is rather easy to double-spend them, but of course this is not allowed. Thus the predefined rule is that a coin can only be spent once. In the system that Chaum designed it is not possible to trace a coin back to its owner if it is spent at most one time, however when a coin is spent more than once, i.e. if the rule is violated, the identity of the owner is revealed.

For this document, we adopt a more liberal interpretation of the original definition of revocable privacy where not only violation of a rule can be the trigger, but also conformance to a rule. The latter viewpoint better matches cases where the user himself conditionally reveals personal information. Furthermore, we also generalize the concept of personal data, to include any type of data that should remain confidential. One example of this type is specific information about company processes or company results. In these cases, it would of course be more precise to talk about revocable confidentiality instead of revocable privacy.

The use cases explored in this report come from various sources. Some of them are real, others are purely hypothetical. In many cases the legality and/or morality of the situation described in the use case are up for debate. We have included them for the sole purpose of investigating the types of rules a system for revocable privacy might need to implement in the future. In no way should inclusion of a use case in this report be interpreted to mean that we endorse the use case in any way.

The structure of this document is as follows. In the next section we briefly discuss the classification we used for the use cases we discovered, before discussing the actual use cases. Finally, we summarize our results.

2 Classification

In this section we describe a classification of the use cases and the basis for our privacy sensitivity analysis.

2.1 Use cases

To prevent obtaining a large list of use cases we classify the use cases based on the underlying (predefined) rules that govern the release of identifiable information, while also taking into account the number of information sources in the rule.

The first two classes contain the simplest rules. The first class contains rules that are primarily threshold based. Such rules apply if some event happens too often, or not often enough, in which cases anonymity is revoked. The electronic cash system of David Chaum [2] is an example of such a system. The second class contains rules that are time based. These rules stipulate that if some event (or sequence of events) takes too long, or too short, this will lead to revocation of anonymity.

The next class of use cases is a bit more complex. We consider completely arbitrary rules, which we call predicates, but limit the data to come from a single source. If the predicate matches, identifying information is revealed. This class is complemented by the class of complex rules, where the restriction on the data sources no longer applies.

Finally, we have a special class of use cases: those with decision based rules. Here a critical part of the rule is that a specific entity, or group of entities has to give permission before anonymity is revoked. In our use cases these entities are people, but this does not necessarily have to be the case.

2.2 Privacy Sensitivity

For every use case we also give an assessment of the privacy sensitivity of the case. This rough assessment tries to quantify the impact of releasing the information that would be stored for the naive solution, where all data is stored and the rules are applied afterwards. This quantification is based on the following factors, which are similar to the Dutch law on privacy protection [3]: accessibility, sensitivity, level of detail and number of people affected.

If the data stored can easily be obtained by any civilian the cost of releasing is lower than if almost nobody can obtain this data. Next we consider the sensitivity of releasing the data; releasing somebody's medical information is deemed more costly than somebody's diet. The final two aspects, amount of data and scope of collection are similar in nature. Collecting only a license plate of a car is much less invasive than taking detailed photographs of the car and its passengers. Also, monitoring a single street in the city centre is much less invasive than monitoring all the streets in a county.

We emphasize that this quantification should be considered only as an indication of the potential impact; a thorough impact assessment is beyond the scope of this document.

3 Use cases

3.1 Threshold based

3.1.1 Canvas cutters

Description. One of the KLPD[1] use cases is the following. Trucks parked at rest stops are interesting targets for criminals to steal goods from. They do this by quickly cutting the canvas that covers the goods in those trucks. One way to detect these criminals is to record which cars frequently enter several rest stops in sequence. Apart from identifying potential criminals this will also identify police cars and vehicles offering roadside assistance, but in general not ordinary cars.

The traditional solution to this problem is to set up ANPR (Automatic Number Plate Recognition) systems, store all license plates that enter the rest stops, and count how often they do so. This means that all cars entering the rest stops are registered and this information can later be queried. A more privacy friendly approach would be to build a system that stores all this data in such a way that only those cars that visit multiple rest stops are revealed.

Privacy sensitivity. Medium

Abstract rule. Consider a system consisting of: a set of entities with unique identifiers that can generate a specific event, a set of parties that can register such events and a threshold. The system will only output an identifier when it has been registered at least as often as the threshold.

Source. Discussion at KLPD.

3.1.2 Anonymous preselection

Description. Consider the following situation. A committee, of for example 20 people, wants to select a chair from within this committee. To do so, they want to make a preselection of candidates that have significant support. Normally, the committee would hold a public vote for the short-list and only include those that passed a certain threshold. The downside to this approach, as with all voting, is that all vote counts are public. Hence, when a person receives few votes a lot of information is can be deduced.

So we propose the following more privacy-friendly approach. All members vote, but only those names that actually occurred often enough will be revealed.

Privacy sensitivity. Low

Abstract rule. Given a set of parties, a set of items and a threshold, each party will select one item. At the end, only those items will be revealed that were selected more often than the predefined threshold. Same rule as 3.1.1.

Source. Internal discussions and brainstorms (hypothetical use case).

3.1.3 Social welfare fraud

Description. In the municipality of Groningen a new system is used to detect welfare fraud. One possible way to commit fraud is to claim that you live alone, while you

[1] The KLPD (Korps Landelijke Politiediensten), is the Dutch national police force.

actually live together with someone else. This is beneficial, because when you live alone you receive higher benefits. In Groningen these cases are detected by combining records at the social welfare organization UWV[2] with data from other sources (see [4]):

- public housing registers, to verify that someone actually lives at a certain address;
- water and energy usage, to verify that water and energy usage correspond to the number of people that are supposedly living at a certain address;
- Internet (usage) and (sometimes) waste disposal information, for the same purpose

Note that none of these sources can normally be directly accessed by the UWV. Any naive implementation for combining these databases will result in considerable leakage of privacy sensitive information: (1) because external sources learn who receives social welfare benefits, and (2) because the social welfare organization also obtains information on people not receiving social welfare.

From the preceding we conclude that any system that implements this functionality, sidestepping for now the question whether it should exist at all, must adhere to the following basic privacy constraints: data should not leave its original context, except for the small part the UWV wants to learn. More formally: none of the external parties learn anything new about their customers by using this system, and the UWV only learns of the people receiving benefits who use significantly fewer resources than expected.

Privacy sensitivity. High

Abstract rule. Consider a set of objects with unique identifiers, and a set of databases containing information about these objects. Furthermore, there is a set of predicates, one for each database, that only uses data from the corresponding database. The goal of this system is to reveal the identifiers of those objects that satisfy the predicates for all (or pre-specified subsets) of these databases.

Alternative settings and extensions. As an extension one could incorporate student benefits. In the Netherlands one is not eligible for social welfare when also receiving student benefits [5]. For this reason municipalities are now allowed [6] to check whether applicants for social welfare also receive student benefits.

Source. Project member discussion and brainstorm.

3.1.4 Object surveillance/security

Description. Where canvas cutters can be detected using the fact that they show up at multiple parking lots during a short period of time, we can also try to detect terrorist groups that are planning attacks on say a nuclear power station. One possible indicator would be that a member from such a group would scout the location a number of times, before commencing the final attack.

To detect such a person a couple of ANPR systems can be placed that will monitor all traffic and then detect vehicles that show up frequently around the installation. This case has a number of subtle differences with the original canvas-cutters scenario:

- First, a car that is seen twice by the same ANPR camera should still be counted twice, and not only once as it is in the canvas-cutters scenario.

[2] The UWV (Dutch: Uitvoeringsinstituut Werknemersverzekeringen) is, amongst others, responsible for providing unemployment benefits and social welfare.

- Second, such a system will lead to false positives — much more so than in the canvas-cutters case. In fact, almost every car that is detected in the neighbourhood of a nuclear power plant will belong to an employee or a vendor. So care has to be taken that these cars are never registered by the system, lest the privacy of the employees and vendors is violated.

Privacy sensitivity. Low, depending on the object under surveillance.

Abstract rule. Consider a system with a set of entities with unique identifiers that can generate a specific event, a set of parties that can generate such events and a threshold. The system will only output an identifier when it has been registered at least as often as the threshold.

Alternative settings. The specific object in this use case does not really matter, as long as it is possible to obtain the unique identifier for an entity that generates an event.

Source. Discussion at KLPD.

3.1.5 Detection and registration of child abuse

Description. Detecting child abuse is difficult due to the fragmented nature of the observing parties: children (and/or their parents) come in contact with a wide variety of organisations and government institutions, e.g., health-care workers, social workers, and teachers. All of these may pick up on some signals pointing to child abuse, but at best they only have a partial picture. One solution to this problem would be to use a central database to store reports of indicators for possible child abuse, but such a database has downsides:

- The threshold to file a report can be (too) high. For example, a doctor might refrain from filing a report, knowing that when he is mistaken there will be a lot of trouble for both the parents and the kid.
- The information in such a database is very privacy sensitive. If there is a way to search this database for example, then this would make it possible to blackmail the parents.

In a revocable-privacy inspired system the identity of a child will only be revealed when there are enough indicators for child abuse. If these indicators are not present it is not possible to learn anything about the child in question.

Privacy sensitivity. High

Abstract rule. Consider a system with a set of entities with unique identifiers that can generate a specific event, a set of parties that can generate such events and a threshold. The system will only output an identifier when it has been registered at least as often as the threshold. This is the same rule as for use case 3.1.4.

Alternative settings and extensions. Just like child abuse also spousal abuse and domestic violence in general match this setting where information sources are scattered and both the ‘cost’ of a false positive as well as a false negative is very high.

Source. Project member discussion and brainstorm.

3.1.6 Anonymous waste disposal

Description. In some cities an electronic system is used to register the amount of waste that is offered for disposal (see also use case 3.1.4). Common methods are a barcode on the disposal bin or a subterranean disposal facility that can be accessed using a

smart card. In both cases it is technically easy to log specific times, sources and amounts of waste disposal.

How much information actually needs to be stored heavily depends on the billing process of the relevant government body. When the exact amount is billed we have similar considerations as the “Road Pricing” use case later on. However, when there is a fixed fee up to a certain maximum a more privacy friendly solution is possible. Only when a user exceeds the maximum does the municipality need to know. The website of the municipality Groningen [4] suggests that they also have access to the amount of waste that households dispose, while they employ a fixed-fee system.

Note that this use case is not the same as the canvas cutters use case (see 3.1.1) because people usually will use the same disposal site whereas in the canvas cutters use case we explicitly count only *different* stations. This also leads to differences in the achievable levels of privacy.

Privacy sensitivity. Low

Abstract rule. Given a set of entities with unique identifiers that can generate a specific event, a set of parties that can generate such events and a threshold, the system will only output an identifier when it has been registered at least as often as the threshold. Same rule as 3.1.4.

Alternative settings. Note that a similar type of use case would be to deal with data limits and fair use.

Source. Project member discussion and brainstorm (hypothetical use case).

3.1.7 No-shows in anonymous reservation systems

Description. Suppose people with an unrestricted subscription to see movies at the local movie theatre can opt to book some seat in this theatre for themselves and friends. This way they are guaranteed that when they show up they will have seats. For this particular purpose it is not necessary for the movie theatre to know who booked, as long as you show up (the tickets are checked at the gate and can be independent of the booking). However, seats that are booked should also be occupied, so if you frequently do not show up, without cancelling your reservation, you should be banned from making any further bookings.

In a traditional system this is easy to implement since we can count how often a person books seats and how often he/she does not show up. When booking is anonymous we need a different primitive. One method would be to use a threshold based scheme, if a person does not show up say three times, his/her identity is revealed and we can block this person from making any further bookings. However, one could also make the system a bit more lenient and only disallow further bookings from this person, without revealing his/her actual identity.

Note that in both cases it is difficult to actually handle the not showing up event in a secure manner. It is in general very easy to modify the system to forget that a person did show up, and hence use this to obtain information about the user.

Privacy sensitivity. Depends on application.

Abstract rule. Consider a set of entities with unique identifiers and a set of objects. An entity can generate a claim event and a release event for a given time slot. The general rule is that if an entity claims an object it should also release it. If said entity fails to issue the release event its identity is revealed.

A more general approach would be to allow an entity a limited number of non-release time slots before its identity is revealed. Instead of releasing the identifying information the system could also block the violating entity from making more claim events.

Alternative settings. This use case is not limited to the movie theatre alone. Any settings where people can book things in advance (without also paying for them directly) is suitable for this approach. For example, a similar system might be used to book multiple seats on a train (as is possible in Germany).

Source. Project member discussion and brainstorm (hypothetical use case).

3.1.8 Fuzzy monitoring

Description. Camera surveillance is becoming increasingly popular. However, at the same time, storing all this footage causes a privacy risk. One way to mitigate this risk is to specifically tailor the data to be stored to the specific task at hand. The problem with camera surveillance, and in particular for the surveillance of humans, is that, generally, we are not able to convert an image into something simple like a social security number. We have to make do with other derived (fuzzy) features which we can match later.

A first example is a jewellery shop. It is well known that robbers scout potential jewelleries up to a couple of times, possibly without even entering the shop, before attempting a robbery. So we want to detect people that frequently scout the shop. One may recognize a simple threshold scheme here, but with the added difficulty of identifying the same person between occurrences. Similar detection mechanisms might also be useful for other cases such as detecting possible terrorists.

Another example is to detect people that stay at the same general location for a long time, where in general this is not normal behaviour. For example somebody who remains in the vicinity of a government building for a longer period of time. Here we have a long time threshold, but now with the added difficulty of identifying a person from different angles in a consistent manner.

Privacy sensitivity. Medium

Abstract rule. By nature of the use case this is a meta rule. Consider a set of entities that have a unique identifier. These identifiers are not uniquely observable, but only partially and dynamically, i.e., each time the entity is observed the observed value can change. The additional goal of the system is then to deal reliably with this ever changing view of the identifier, as if it witnessed the same, unique, identifier every time the same entity presents itself.

When the dynamically recognized entity satisfies the ground rule two things can happen. First, only the recovered information about the entity can be revealed. Depending on the quality of the conversion process this suffices. Alternatively, satisfying the ground rule can trigger the release of the original observed (partial) information.

Source. Project member discussion and brainstorm (hypothetical use case).

3.2 Time based

3.2.1 Drug runners

Description. Because the drugs policy in the Netherlands is more liberal than in neighbouring countries, border towns suffer from excessive drug trafficking. One of these border towns is Maastricht, at the border with Belgium and Germany. Typically,

large amounts of drugs are imported from Rotterdam. A typical pattern for these drugs runners is that they make trips from Maastricht, to Rotterdam and back again in a short period of time.

To detect these drug-runners one could strategically place a couple of ANPR systems along the main highways and simply log all the traffic. A more privacy friendly approach would be to only detect cars matching the specific criteria, i.e., they travel to and from Rotterdam within a short period of time.

Privacy sensitivity. Low-Medium

Abstract rule. Consider a set of entities with unique identifiers, a predefined sequence of events, and a time limit. Over time, entities generate a sequence of events. We say that such a sequence matches the predefined sequence if the latter is a subsequence of the former. The goal of the system is to reveal the identifiers of those objects that match the predefined sequence and the entire predefined sequence occurs within the time limit.

Source. Discussion at KLPD.

3.2.2 Average speed checking

Description. The most commonly used techniques for speed limit enforcement nowadays is the use of laser guns and radar. These only provide a measurement for a single point in time. In, for example, the Netherlands an alternative system is in use that measures the average speed of a car along a fixed-length trajectory.

These systems use two ANPR cameras, one at the start of the trajectory, and one at the end. When a car passes the first ANPR camera its license plate and the current time is stored. On passing the second camera the system will look up the license plate and the corresponding start time and uses the latter to determine the average speed. All cars exceeding the speed limit are then stored.

The privacy friendliness of this approach relies on the quality of the implementation. If it is possible to later recover (some of) the stored license plates that did adhere to the speed limit the system is not privacy friendly.

Privacy sensitivity. Medium

Abstract rule. Given a set of entities with corresponding unique identifiers, start and stop events, and a time limit, the goal of the system is to reveal the unique identifiers of those entities that issue a start and subsequent stop event within the time limit.

Source. Project member discussion and brainstorm (hypothetical use case).

3.2.3 Time limited resources

Description. This hypothetical use case concerns power usage at a camping. To discourage people from using too much power, a fine is charged when they use more than 100W of power for longer than 15 minutes. One method for solving this is to record power usage every minute and then calculate the costs afterwards. However, then very detailed power usage information is stored, which is possibly privacy invasive. It would be better to only store those people actually deserving the fine.

Privacy sensitivity. Low

Abstract rule. Given a set of entities with corresponding unique identifiers, start and stop events, and a time limit, the goal of the system is to reveal the unique identifiers of

those entities where the time between a start and the subsequent stop event exceeds the time limit.

Source. Project member discussion and brainstorm (hypothetical use case).

3.3 Predicate based

3.3.1 Forbidden unique identifiers

Description. The Dutch wiretapping system allows law enforcement officers to request access to the complete data stream of a suspect's Internet connection. Investigators can then perform all kinds of analysis on this data. Sometimes criminal or suspicious behaviour is almost uniquely identifiable from a single action and can hence be easily detected. One of these actions would be a computer opening a URL that is known to contain child pornography. Similarly, one could trigger on the hash of a known bad image. Finally, also the action of connecting to (or even issuing commands to) a botnet root host could lead to a trigger. So to decrease the invasiveness of all-out wiretapping we could instead focus on these events.

Traditionally, the complete data stream is available at the ISP, so they could take care of extracting the identifiers and comparing them to suitable blacklists. However, this also means that the ISP learns which customers exhibit possibly criminal behaviour, which is not desirable from a privacy point of view. Sending all the identifiers to the law enforcement officers would, on the other hand, still reveal more information about the suspect than is strictly necessary.

We propose the following alternative. The ISP receives an encoded blacklist and encodes each identifier and corresponding meta information against this blacklist. The result is sent to the law enforcement officer. The system will only reveal those identifiers, together with the corresponding meta information, when the identifier was actually on the blacklist. Note that the ISP learns nothing of this result, while the officer only learns about violating entries.

We have to emphasize that this kind of deep packet inspection is only allowed under the wiretapping law and never for any general set of civilians. Hence, we see this approach more as an alternative to traditional wiretapping, than as a way to monitor everybody.

Privacy sensitivity. Medium-High

Abstract rule. Given two parties, a logger and an auditor. The auditor produces a list of suspicious identifiers. For each event the logger encounters it stores its identifier in the system together with the relevant meta data. The system will reveal to the auditor only those identifiers and corresponding meta data entries that match the list. The logger learns nothing.

Alternative settings. Instead of focusing the system on a single suspect one could increase privacy further by including the identity of a suspect in the meta information and then just send all encoded information to the law officers without revealing the corresponding party. In this way the possibility for traffic analysis would be further reduced.

Source. Project member discussion and brainstorm.

3.3.2 *Automatic scanning of e-mail*

Description. Big companies/organizations often have security officers that are in special circumstances allowed to look through other employees' email to detect malicious behaviour. These officers are bound by strict rules describing when and how they are allowed to examine these emails. We can make a more privacy friendly version of this system by

1. encoding the rules for the officers such that it is not possible to violate them, a variant of this can be seen in the wiretapping use case, where the system ensures that if the officers must have previously found some evidence this is actually checked; or
2. a precise encoding of the criteria by which the e-mail is examined. Once we have this encoding, the email client extract some meta-data from each e-mail that is sent, and transmits this for central storage. This meta-data allows the security officer to check against the criteria once they have a need to do so. Only the matching emails are revealed.

This type of approach primarily increases privacy by taking the human out of the loop. This system is an example of a predicate on data. Note that this use case is a generalization of the previous use case as we now allow any predicate.

Privacy sensitivity. Medium--High

Abstract rule. Given a set of objects and a precisely defined predicate over these objects. Furthermore, consider a logger and an auditor. The auditor only learns of those objects administrated by the logger that actually match the predicate.

Alternative settings. This approach would also work in the wiretapping setting from the previous use case.

Source. Discussion with Richard Kerkdijk (TNO).

3.3.3 *Generalized forbidden image detection*

Description. The downside of current forbidden image detection systems is that they are mostly hash-based, see also the unique identifier case. This has some serious disadvantages: (1) This system requires explicit classification of forbidden images, which is time consuming to maintain; (2) it does not work for new images; and (3) it is very easy to cheat the system by making minimal changes to the image. As an alternative it may be possible to make a general function that will evaluate an image and determine whether it contains some forbidden content. By making this a general function the disadvantages of a hash-based systems are mitigated.

Just as in the previous use cases this function can be applied at the ISP, without revealing the result. A law enforcement officer can then later, given the right permissions, see only which forbidden content a suspect accesses. The privacy concerns we raised for the automatic email scanning use case and the forbidden identifier use case also apply here.

Privacy sensitivity. Medium-High

Abstract rule. Given a set of objects and a precisely defined predicate over these objects. Furthermore, consider a logger and an auditor. The auditor only learns of those objects administrated by the logger that actually match the predicate. This is the same rule as in use case 3.3.2, however, most likely the predicates will be much more complex here.

Alternative settings. This setting is not limited to static images alone. This type of encoded image processing software could also be applied to privately owned video camera systems to detect criminals and terrorists, without the owner of the camera learning about this.

Source. Project member discussion and brainstorm.

3.3.4 Secret rules

Description. This is a hypothetical use case. Consider a bureaucratic process at a government institution that (1) requires the transfer of personal information of a civilian to the institution and (2) based on that information a decision is made whether the civilian gets permission of some sort. It is in the interest of the civilian to get this permission, so it is willing the share his or her information. However, if the rules for the decision are not public or very complex, it becomes infeasible for the civilian to check whether it is worth it to apply for the permission.

Traditionally the civilian would supply his/her personal information and ask the government institution to determine whether it is eligible to get the permission. While a government institution may be trustworthy, not every institution is, nor will every civilian be as willing to part with highly personal information. Hence, it would be nicer to have a system where the personal information is only revealed if the condition for getting the permission is met. Additionally, in some cases the rules might actually be secret, thus increasing the need for a privacy friendly solution even more.

Alternatively, one could also look at the inverse. In this case this means that (some of) the identity/information is revealed when the rules are *not* met, making it possible for the institution to give feedback to the civilian. At the same time, the cases that do match the rules get the permission anonymously.

Privacy sensitivity. Medium

Abstract rule. Given a predefined predicate on a record, an entity can interact with the system as follows. It provides its record to the system, only when the record matches the predicate will the record be revealed. Alternatively, the system will either indicate that the record matches the predicate, or reveal why the record does not satisfy the predicate.

Source. Discussion with Rieks Joosten (TNO).

3.3.5 Privacy friendly matching

Description. This is a hypothetical use case. Consider a social network that is designed to let people meet new and interesting people. One method for doing this would be for everybody to reveal their interests, sex, etc. You can then select potential friends based on these data. However, this personal information is very privacy sensitive, and it would be preferable if this information is not shared at all.

The only information that really needs to be shared, after two people decide they want to meet, is some identifiable information. This information should only be revealed if the two people match, i.e. the first person likes the second person, while the second likes the first. Important here is that the system should allow people to determine whether they like each other, without explicitly revealing the underlying data.

Privacy sensitivity. Medium

Abstract rule. Consider two parties, each with a unique identifier, a number of records and a predicate on these records. The system will reveal the unique identifier to the other party if each party's records satisfy the other party's predicate.

Source. Project member discussion and brainstorm.

3.4 Complex rules

3.4.1 Retrieval of camera footage after the crime

Description. Again we consider a surveillance system using cameras, for example in the centre of a city. These cameras are used to monitor and detect violence and criminal behaviour. Generally, it would suffice not to store any of these images, and only look at them in real time like a normal officer would. However, once a crime has been detected it is usually worthwhile to be able to go back and see what happened before and to store this for later reference.

The simple method of implementing this rewind functionality would be to always store the last 2 hours. This is, however, much less privacy friendly than only showing (and not storing!) the current image, as now the previous two hours are always accessible. We could increase privacy by formally defining what constitutes a violent or criminal situation, and let the system decide when to release previous information.

We note that this use case is not the same as in the wiretapping cases as there a person will actually decide to reveal the data, while here this happens automatically, using the proof that there was criminal or violent behaviour. This use case is therefore more along the lines of the automatic e-mail scanning and generalized forbidden image detection cases. However, where in those cases we consider individual data objects, we here consider a stream of data objects where a condition in the future determines the match.

Privacy sensitivity. Medium

Abstract rule. Consider a sequence of data blocks, together with a selector predicate and a combiner relation on these data blocks. The system will process the data blocks in sequence. If it finds a data block that matches the selector predicate it will output not only this data block, but also a continuous sequence of previous data blocks up to the first data block that does not satisfy the combiner relation with the first.

Source. Discussion with Richard Kerkdijk (TNO).

3.4.2 Terrorist detection with multiple clues

Description. Contrary to the canvas cutters use case a lot of law-enforcement-like cases depend on combining various indicators to find the bad guys. One, rather primitive, example works as follows. A person that buys fertilizer, rents a van and scouts a government building in a short period of time may be planning to make and set off a bomb.

More abstractly we have a number of events here (possibly within separate scopes), which we combine into one global indicator using logical connectives. We could extend the previous by requiring that this person is not a farmer or that instead of only fertilizer we also allow chemical equipment useful for making explosives. Hence, we also need or-connectives and not-connectives to make the most general system.

Privacy sensitivity. Medium

Abstract rule. Given a set of entities with unique identifiers, a set of sensors that can make claims about an entity and a predefined logical formula on these claims. When a sensor encounters an entity they can issue claims on this entity to the system. The system will only reveal the unique identifiers of those entities whose claims match the logical formula.

Alternative settings. As indicated by the abstract rule one is of course not limited to finding terrorists in this specific manner, nor is one limited to finding terrorists at all. This rule is applicable to any system where the sensor output should remain hidden, while we are on the other hand interested in logical formula on sensor data for a specific individual.

Source. Project member discussion and brainstorm (hypothetical use case).

3.4.3 Anomaly detection

Description. The following use case deals with eggs and the detection of fraud with these eggs. Depending on how the chickens are kept a different classification is assigned to their eggs. For example, the eggs of free range chickens get a better classification than those of chickens in battery cages. Eggs with a better rating fetch higher prices. Therefore, it is beneficial for say a farm with free range chickens to obtain some additional eggs from battery cages (at a low price) and subsequently sell them at a higher price as free range eggs.

To prevent this type of fraud the Netherlands Voedsel en Warenautoriteit (VWa) tries to tally the total number of eggs sold by every farm (they are allowed to sell to different distribution centers). To protect the business interests of the farms it should remain hidden how much they sell to each distribution center. On the other hand the total number of eggs sold each weak should be compared against the record known at the VWa.

This is an example of a use case that is not really oriented towards privacy, but instead has a focus on business processes in which participants want to detect a global ‘event’, but at the same time simply sharing all information to achieve this is not an option because of the sensitive nature of the information.

Business risk: Medium

Abstract rule. Consider a set of entities with unique identifiers. Each of these entities generate a total number of events that is known to the system, finally, we have a margin of error. Every entity can generate events at a number of sensors. At the end of a time slot the entities will report the total number of events of each entity to the system. The system will reveal the unique identifiers of those entities whose total number of events differs more than the margin of error from the expected total.

Source. Discussion with Jan Pieter Wijbenga (TNO).

3.4.4 Riot control (people flow)

Description. During the turn of the year 2009/2010 there were riots between two ethnic groups (Moluccan and Moroccan) in Culemborg. The police was informed that the rioting groups were getting ‘help’ in the form of friends from other parts of the country (mainly the the Randstad). To safeguard public order, it was important for the police to know if and when such ‘reinforcements’ were on their way to Culemborg. It turned out that such ‘friends’ were mostly from some specific regions, which were known by their postal code.

To prevent the ‘reinforcements’ from reaching Culemborg the police used a system where cars with a mobile ANPR-system patrolled the highways and scanned all license-plates. Then, if more than four vehicles registered to the special set of postal codes were detected within a short period of time, a warning signal was issued, triggering a temporary blockage of the highway exit to Culemborg. For checking the postal codes of the vehicles the police used a real-time link to the RDW.

This use-case is an example of a more complex rule, one where there is a link to another (independent) database. Also, there is the issue of the RDW now knowing, based on the queries from the police, the location and time of all the vehicles scanned. Finally, all license plates have to be stored for some time, which is also a bit privacy unfriendly.

Privacy sensitivity. Medium

Abstract rule. Consider a set of entities, each with a unique identifier, a set of (possibly external) databases that can (indirectly) provide information on these entities, a predicate over the data corresponding to an entity in these databases, a threshold and a time limit. The system observes the entities, but will only report them if the number of entities that matches the predicate exceeds the threshold within the given time limit.

Source. Discussion at KLPD.

3.5 Decision events

3.5.1 Road pricing

Description. One of the ways to battle traffic jams is to simply build more roads. However, building and maintaining roads is expensive, and for this reason the government is starting projects to make the actuals users of the roads pay (more) for using these roads. To be able to correctly bill the users a system would need to track and register the cars on the roads. Storing all these car movements in a database would be potentially privacy invasive as the information in such a database would pinpoint a person (or at least the car) to a certain position and time.

At the end of the month the system needs to be able to bill the customer, so it should know the accumulated cost. However, individual trips do not necessarily need to be stored. On the other hand, in this way it is never possible for a car owner to dispute a bill, because in order to do so the system needs to show the individual tracks. A privacy friendly solution would allow the system to reveal these tracks, only when the corresponding car owner requests this, i.e., he/she decides to reveal the travel information.

Privacy sensitivity. Medium

Abstract rule. Given a set of entities with unique identifiers, and a number of sensors that can produce measurements on these entities. When an sensor witnesses an entity it submits a measurement to the system. At the end of the time frame the system outputs the sum of all these measurements. The individual measurements are only revealed when the corresponding entity requests this.

Alternative settings. This approach could be used for any type of verifiable tally, where normally only the total matters. Other examples include credit cards and public transport system cards.

Source. Project member discussion and brainstorm (hypothetical use case).

3.5.2 *Anonymous medical data sharing with feedback*

Description. An important part of medical research is obtaining (large) amounts of data for analysis. As such, this depends on the willingness of people to share potentially privacy sensitive information. Therefore, the data is usually anonymized, such that (hopefully) it is not possible to link individuals to specific data entries. However, one can imagine that in some cases, in particular in the medical domain, it is sometimes in the best interest of the individual to be identifiable, for example when the researcher discovers that the individual in question may suffer from a serious illness.

Hence, we would like a system that allows individuals to share information anonymously, but at the same time can be contacted if something comes up they should be notified of. One example of such a system was proposed by Galindo and Verheul [7], but here only the original supplier of the data has the power to revoke the anonymity. Ideally, however, this decision is not made by the data supplier, but is instead automatic according to a rule (pre-)defined by the patient.

At this point it is not clear whether it is possible to define these rules up front, especially not because ongoing research may reveal new methods. Moreover, if the rules were known up front, then they could conceivably be applied by the patients themselves, forgoing the need for a researcher all together. It may be the case that interaction (in some anonymized manner) with the patient is necessary such that the patient can determine the significance of the find.

Privacy sensitivity. High

Abstract rule. Given a set of entities with unique identifiers, and data on some of these entities, and a researcher. Each entity defines a predicate, which is allowed to depend on other data entries. The research is given the data on the entities in anonymized form. The system will only reveal the unique identifier of an entity to the research if the predicate is satisfied.

Note that the generality of the predicates in this case makes it different from other use cases. Here we allow the predicate “my haemoglobin levels significantly differ from those of `other patients”, which would normally not be allowed.

Source. Project member discussion and brainstorm (hypothetical use case).

3.5.3 *Wiretapping*

Description. In the Netherlands wiretaps for phone and Internet can be enabled centrally. By law either a prosecutor (Dutch: officier van justitie) or a judge (in the Netherlands this will then be the “rechter commissaris”) has to sign off on the placement of such a wiretap. The requests for tapping can be send digitally, where the permission by the required authorities is assumed to be correct. Whether permission was really given can be checked via other channels.

To increase security of this system, and hence privacy of ordinary citizens, a better verification of these permissions is necessary. In fact, ideally the system would enforce that no wiretapped data is released to the law enforcement officers, unless a valid and verifiable request has been received. This request would then serve to revoke the privacy of the person who is being tapped.

An additional difficulty is that sometimes it is not possible to wait for authorization to come through (for example in case of a hostage situation or terrorist attack). In this case the system should provide the data only when very strong access logging takes place and formal auditing happens afterward.

Privacy sensitivity. Medium

Abstract rule. Consider a system that can monitor the data streams produced by a set of entities. Furthermore consider a predefined policy describing when an auditor is allowed access to the data stream. The system will only provide the auditor with the data stream corresponding to an entity after a valid request, i.e., a request that satisfies the policy, is received.

Source. Discussion with Frank Fransen (TNO).

3.5.4 Group decision for anonymity revocation

Description. Most of the previous cases dealt with revocation of anonymity either as a result of directly violating a rule, or a person deciding that a rule was violated. In this case this decision has to be made by a group of people. Consider an online community that, for privacy reasons, has decided to give users full anonymity (how exactly to achieve this is beyond the scope of this report). On the other hand, the community realizes that full privacy would make it a safe haven for behaviour that they would rather not or could not accept. Hence they build in a safeguard that allows other parties to request the revocation of anonymity. Only when a predefined number of members agrees with the need for revocation can the anonymity be revealed.

While this is a powerful idea one has to take care of a number of important issues:

- How does the system cope with dynamic group sizes. Does the limit change when the group size changes?
- And if the group is dynamic, do only the members at the time of posting (or maybe even joining) vote to revoke anonymity, or can also newer members do so?

Privacy sensitivity. Depends on content.

Abstract rule. Consider a system with a set of entities with unique identifiers, and a predefined limit. Each entity can anonymously publish blocks of data. Upon request the system will reveal the identity of a poster of a block of data, but only if more than the predefined limit of entities vote to do so.

Source. Discussion with Michiel Prins (Online 24).

4 Summary and future work

In this document we have described and partially analysed a number of use cases where revocable privacy can help in providing more privacy for the user. We have seen that the use cases can be classified based primarily on the type of rule they encode. We classified threshold based, time based and predicate based rules. In addition we considered decision based rules that involve an explicit decision component and finally a set of complex use cases that combine some of the previous aspects.

For future work we believe that there are a couple of interesting directions. First, we have uncovered a couple of use cases that should admit a relatively easy implementation which would make nice demonstrators. Second, some of the use cases employ simple primitives that are ideal targets to really design new cryptographic primitives for. Third, the more complex use cases could really benefit from a more thorough analysis in terms of the exact rules that are necessary and a tight description of the functionality that is desired. Finally, our analysis of the privacy sensitivity should be made more rigorous.

5 Bibliography

[1] J.-H. Hoepman, “Revocable Privacy,” *ENISA Quarterly Review,* vol. 5, no. 2, pp. 16-17, June 2009.

[2] D. Chaum, “Blind Signatures for Untraceable Payments,” in *Crypto*, Santa Barabara, Calfornia, U.S.A., 1982.

[3] *Wet bescherming persoonsgegevens,* 2002.

[4] Gemeente Groningen, “Sociale dienst spoort bijna driehonderd gevallen van bijstandsfraude op,” 29 01 2010. [Online]. Available: http://gemeente.groningen.nl/algemeen-nieuws/2010-1/sociale-dienst-spoort-bijna-driehonderd-gevallen-van-bijstandsfraude-op. [Accessed 05 01 2012].

[5] Schulink, “Recht op Bijstand - U studeert,” 2012. [Online]. Available: http://www.rechtopbijstand.nl/inhoud?pid=23. [Accessed 05 01 2012].

[6] ANP, “Aanpak bijstandsfraude door bestandskoppeling,” 19 11 2011. [Online]. Available: http://www.nu.nl/politiek/2670044/aanpak-bijstandsfraude-bestandskoppeling.html. [Accessed 05 01 2012].

[7] D. Galindo and E. R. Verheul, “Pseudonymized Data Sharing,” in *Privacy and Anonymity and Knowledge Processing*, J. Nin and J. Herranz, Eds., London, Springer, 2010, pp. 157-179.