Privacy Analysis of a Mobile Sensor Application

Heinrich Hartmann, Tim Wambach, Maximilian Meffert, and Rüdiger Grimm

University of Koblenz-Landau

Abstract. TODO: (later)

Keywords: privacy protection, IT security analysis, sensor data, mobile phones, traffic survey

1 Introduction

TODO: RG
* Context: Reference model IT Security analysis
* Application: Travel Survey
* Different definitions of privacy in the literature.
* Study outcomes depend on privacy definition

In recent days the importance of privacy protection has been amplified by the reports about the mass surveillance of ordinary citizens on a global scale by the NSA and other intelligence agencies around the world.

While aiming at the noble cause of enhancing eParticipation using mobile technologies, Live+Gov systems do process a large variety data that is potentially infringing the citizens privacy. The captured data includes personal information like name, phone numbers and email addresses and sensor data from GPS and accelerometer sensors. Also with some applications it is possible to gather images and textual input from the citizen.

While the collection of this data is necessary for providing the advanced services that Live+Gov aims to deliver, at he same time, the available raw data can be used to draw a very detailed picture of the private life of the citizen. For instance can GPS location tracking be used to reveal shopping habits (e.g. when a car seller is visited) and associations to political groups (when a meeting is attended). Accelerometer data can be used to infer medical conditions like walking disabilities. Images can contain faces of nearby persons to with whom the citizen is associated. All this data is highly sensitive to the citizens privacy and can be used against the citizen if it falls in the wrong hands.

The great importance of protecting the citizens privacy should be apparent from these examples. The European Union, as well as many other countries in the past, has set out a number of directives that regulate the collection, processing and use of privacy sensitive data. We explain the most relevant legislation in Section **??**.

The ethical aspects of privacy have been the subject of study of many social scientists and philosophers. One scholar which is particularly relevant in our context is Charles Fried. He investiages, why we are intuitively so sensitive to violations of our privacy. For him privacy is not asserted as an intrinsic value by itself, he rather sated:

> *Privacy is not simply an absence of information about us in the minds of others; rather it is the control we have over information about ourselves.*

Fried's study on the understanding of privacy provided a great contribution to the research on the same term in philosophy and computer science and despite the fact his text was published in 1970, he already included technologies to its viewpoint (like location monitoring) that are particularly relevant to our Context. We explain his theory in **??** *and follow his definition of privacy in this document.*

In this document we perform a thorough analysis of how to protect the privacy of a citizen and present several implementations that form the core of our privacy aware Sensor Data Storage and Mining Infrastructure.

We identify six main threats for the citizens privacy, and derive eight recommendations that should be followed in order to reduce the associated risks for hazards. This analysis form the guideline for the selection of 9 measures that were implemented in our system and documented in Chapter **??**.

2 Scenario Description: Mobile Traffic Survey

```
TODO: HH:
* System description: (Raw version copied from Privacy Analysis)
* Information needs of transport agency
```

2.1 A Privacy Respecting Mobile Traffic Survey

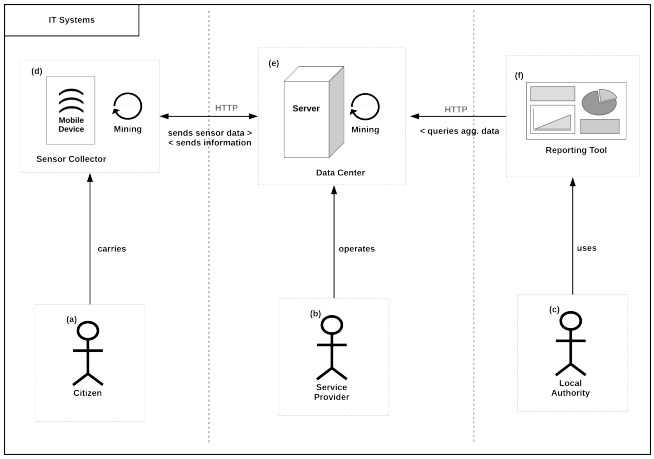
The scenario of our privacy analysis is a mobile traffic survey that supports local travel services by collecting and evaluating mobile sensor data of their customers. The survey is executed by a trustworthy third-party service. The travel services need the personal mobile data for two reasons: firstly, they want to serve their customers individually, for example by guiding them to proximity stations, to recommend adequate travel connections, to inform about delays, or to sell electronic tickets. For this purpose, the personal data for evaluation are related to the individual customers. Secondly, the travel services want to enhance their system, in that they evaluate the demand of stations, connections, changes, ways and time spent by their customers. For this purpose, personal data are anonymized before evaluation.

2.2 IT-Systems & Interactions

This section outlines the general architecture (Figure 1) of IT systems for public monitoring comparable to the Live+Gov project. This includes a description of it technical infrastructure and the interactions between its components.

The IT infrastructure of the Live+Gov system, i.e. the Live+Gov toolkit and the customization of the software components are described in detail in various project deliverables: D4.1, D4.3, D1.1, D5.1. In this section we give an abstraction of those systems from the perspective of WP1.

A citizen *(a) carries a mobile device running the* sensor collector *application (d). The sensor collector application is able to collect various kinds of sensor data (accelerometer, GPS, GSM, ...) and can sent the raw data back to the* data center *(c). The sensor collector can also be able to perform certain data mining operations.*



Legend:

- **(a) Citizen:** User of the L+G client application whose privacy is at stake.
- **(b) Service Provider:** Provides technical infrastructure.
- **(c) Local Authority:** Provider of the L+G system.
- **(d) Mobile Device:** Runs the L+G client application, produces and stores data sensitive to the users privacy.
- **(e) Data Center:** Runs the L+G services, processes and stores user data.
- **(f) Report Tool:** Interface to aggregated user data.

Fig. 1. IT Systems

Examples of such data mining operations include human activity recognition,the detection of service lines, detection of characters and faces from images.

In the Live+Gov Project mobile devices are used in particular for the following activities:

1. *collection of GPS samples,*
2. *mining of Human activities (HAR) based on accelerometer samples,*
3. *collection of reports consisting of an image, free text and selected categories.*

The data center *stores and processes sensor data collected with the sensor collector application. It can also take into account data obtained from third parties, like the current positions of trains. The data center can sent mining end products (traffic jam reports, bus schedule) back to the mobile device of the citizen.*

In the Live+Gov Project data centers, in particular, perform the following activities:

1. *storage of login credentials, name, and email address for each citizen,*
2. *storage of collected GPS samples (user id, timestamp, GPS location),*
3. *storage of HAR results (user id, timestamp, HAR result),*
4. *detection of service lines based received GPS samples (SLD),*
5. *storage of SLD results (userid, timestamp, SLD result),*
6. *sent back SLD results to mobile device,*
7. *storage of received reports (user id, timestamp, report),*

8. *detection of inherent patterns of the received reports.*

The System Provider *(b) provides and operates technical infrastructure like the Data Center and the* Reporting Tool *(f). The Reporting Tool queries the Data Center for aggregated data to visualize in form of charts and other means suitable to help understanding of monitored citizens.*

Local Authorities (c) use the reporting tool to get information in order to understand citizen movement and improve public services. In such systems, the most valuable information for local authorities is not the raw sensor data, but aggregated views on the mining end products.

In the Live+Gov Project the reporting tools allows, in particular the following queries:

1. *show aggregate information about which routes citizens take starting from a given location,*
2. *show the average waiting time of a citizen for each bus stop,*
3. *show routes where citizens where running to catch a bus,*
4. *show locations of all reports in a given time window,*
5. *visualize detected patterns in the report data,*
6. *view individual reports.*

3 Definition of Privacy

More than one hundred years ago, 1890, the American politician Louis Brandis had specified privacy as the right to be left alone [2]. From the data processing point of view, this right is best expressed by the absence of information about a person in the mind of others. Indeed, this principle of "data minimization" is still fundamental to modern data protection legislation. The modern principles of data protection are codified by many legal systems in different countries, especially in Europe [6], and by the US save harbor principles. However, modern data protection is more than only the absence of personal data. It is based on the personal right on self-determination over information and communication [3]. This is, for example, very well expressed by Fried's definition of 1970:

> Privacy is not simply an absence of information about us in the minds of others; rather it is the control we have over information about ourselves. [8]

In order to realize this requirement in the modern IT world users are provided with user control functions, for example to decide about the access on their data for future use, and to view, modify, and delete unwanted data that had been collected. In its fundamental decision of May 13, 2014, the European Court has convicted Google to accept user demands to delete links to incorrect or irrelevant personal data in the Internet (C-131/12 – Google Spain SL). Such functions would constitute a so-called self-data-protection. It requires users to be highly aware of their rights and how to use them. Complementary, system-data-protection puts the load of data protection enforcement on the data collectors,

ideally even without bothering the users. Typical system data protection functions are the deletion of personal data that are not bound to service purpose, the anonymization of personal data for research, and the abstinence from forwarding personal data to third parties. Together, system-data-protection and self-data-protection are supposed to provide a fair balance between the interests of service providers and their users. In our research work we focus on self-data-protection according to Fried's user control requirement [8]. A system data protection point of view would lead to other results, namely to a set of obligations for service providers. This is subject to further research.

3.1 The Seven Types of Privacy

In Fried's definition of privacy as control over information, the specification of what constitutes such information remains open. There is a vast amount of information that relates to a person and we need to get a better understanding in order to perform a thorough analysis. To this end we use the categorization by Friedewald, Finn and Wright of 2013 called the Seven Types of Privacy [7]. These are an extension to the four types of privacy by Roger Clarke of 1997 [5]. It is important to note, that these categories are not mutually exclusive. For instance a written email is considered personal communication as well as personal data stored on a computer. However, the categories are very helpful for a privacy analysis with a focus on self-data-protection. The seven types of privacy are as follows:

1. Privacy of the Person (with respect to body functions and body characteristics)
2. Privacy of Behavior and Action
3. Privacy of Communication
4. Privacy of Data and Image
5. Privacy of Thoughts and Feelings
6. Privacy of Location and Space
7. Privacy of Association (with persons, e.g. friends, and organizations, e.g. political parties)

3.2 Sensor Data Privacy Impact

Modern mobile devices have a broad collection of sensors. Disclosure or processing of sensor data can impact one's privacy. In this section we identify groups of sensors and their potential impact on a certain type of privacy. Figure 2 qualifies that impact on a simple scale. Privacy of Data and Image is trivially threatened because here sensor data is individual data, a priori. Indirect impact is caused by combining sensor data with additional knowledge. For instance, comparing a contemporary map with locational data can imply behaviour if the position matches a church.

	Pricacy of the Person	Pricacy of Behaviour and Action	Pricacy of Communication	Pricacy of Data and Image	Pricacy of Thoughts and Feelings	Pricacy of Location and Space	Pricacy of Association
GPS Sensor	0.5	0.5	0	1	0	1	0.5
Motion Sensors	1	0.5	0	1	0	0	0
Networking Sensors	0.5	0.5	0	1	0	1	1

0: No Impact, 0.5: Indirect Impact, 1: Direct Impact

Fig. 2. Sensor Data Privacy Impact Matrix

GPS Sensor. The GPS sensor gives the current longitude and latitude, the current global position of the mobile device and its carrier, although there is some artificial inaccuracy within civil use. Therefore, the collection of GPS data violates directly the citizens privacy of Location and Space.

Motion Sensors. Accelerometer, Rotation Vector, Gyroscope and Magnetic field sensor measure the physical movement of the mobile device on all three axes. If the mobile device is carried "normally" its safe to say that those sensors also measure the moments of its carrier. So his privacy is infringed regarding biometric behaviour, i.e. the Privacy of the Person.

Network Sensors. The GSM and WLAN sensors reveal the position of the mobile device and its carrier, when used in connection with external databases. The both sensors give the exact cell or network, the mobile device has registered with at the current moment. Frequent connection to one particular network also reveals association, e.g. university networks.

The Bluetooth sensors record lists of the bluetooth clients in the direct neighbourhood. Since those clients are usually moving, inference of the position is usually not possible. Instead, bluetooth clients carried by a third person may infringe the Privacy of Association.

4 Privacy Analysis

The goal of this chapter is to analyze and identify the threats to personal privacy that are posed by collecting, storing and processing sensor data from mobile phones. We derive concrete privacy protection measures that address the main risks involved with handling such data.

In our analysis we follow the "reference model for IT security analysis" as described in [9]. It supersedes earlier efforts by e.g. [1]. The reference model consists of a model and a procedure. The model organizes a common security terminology in a reasonable and practical way. The procedure describes a method for analyzing the IT system based on that model. The reference model provides

four views: (1) the real world of persons and their assets, (2) the potential world of requirements and threats, (3) measures and plans specified by programs, business models and attack strategies, and, finally, (4) events of running programs, data accesses and performed attacks as well as their defense. In the following sections we apply the proposed procedures of the reference model to our scenario of a travel service that collects personal data from mobile users in order to serve the users and to enhance the service. The service intends to respect the privacy of its users.

In sections 4.1 and 4.2 we apply the first two steps of the reference model [9], which are related to the views on the real world and on the measures and plans. In section 4.3 we apply the third step of the reference model by providing specific privacy recommendations and requirements as a result of the previous analysis that the system must comply with.

4.1 Step 1. World Analysis

The first step is the *world view* where all components are described in their current state. It consists of the following components: `Assets`, `IT-Systems`, `Actors`, `Conflicts of Interests`, `Vulnerabilities`, and `Interactions`.

The relevant `IT-Systems` were already described in Section 2.2. `Interactions` between assets, IT-Systems, humans, and vulnerabilites as partly described in Section 2.2 and will be further analyzed in Section 4.2.

Assets In this scenario we focus our attention to only one asset: The privacy of the citizen. Our definition of privacy is described in detail in Section 3 and introduced in Section 3.1 the seven different types of privacy.

Actors This section describes the human actors previously introduced within the IT system architecture (2.2). Also the additional actor *External* is introduced as a person with no special access rights. Although the text is only a short description, a list of the most important interests is provided for each actor.

Citizen. Citizens are persons who use the mobile device as users of the provided software. Their main motivation for using the software is to gain a higher level of convenience in their daily activities. For example they may have access to real time bus schedules or reports about traffic jams, that help them to avoid long waiting times. Also they generally benefit from improvements of public infrastructure by local authorities, which is triggered by issue reports.

By using the application, citizens are sharing personal information like name and address, as well as data gathered from mobile sensor with the service provider. This data can be exploited in ways that are harmful to the citizen (cf. [4]).

Citizens are interested in: physical wellbeing and health, financial profit, convenience, legitimate use of personal data, non-disclosure of personal data to peers of the citizen, and not being monitored.

External. Externals are persons who do not have privileged access to the IT systems, and are willing to break laws, security constrains and norms in order to promote their interests.

A common interest of an external is financial profit. For example they want to obtain access to critical systems to steal sensitive data or to get the system under their control. Stolen data could simply be sold as is or used for illegitimate purposes, e.g. spam or phishing attacks - or excessive data mining.

In short, externals could be interested in: increase power over citizen, financial profit, political activism, and their social standing.

System Provider. System Providers operate the technical infrastructure (hardware and software) of the IT System. They are private companies and legal persons in their own right, but also employ a number of people with diverging interests including administrators, programmer/developer, and a support manager.

As companies, they are interested in gaining financial profit. The financial success of system providers depends on the task complexity of the maintained infrastructure. The complexity of a task has to be in reasonable bounds, so that system providers can complete it within time, with a satisfying quality.

System Providers are interested in: financial profit, manageable complexity, professional excellence, and good working conditions.

Local Authority. Local authorities are public offices (ministry, agency, department, ...) or other external public entities which act as direct customers of service providers. They purchase a system specialized for their needs. For example a department for urban mobility, orders a system to better understand usage patterns and make improvement to the urban traffic flow. Such systems are investments, and so naturally local authorities are interested in a profitable return, like increased ticket sales.

Local authorities are interested in: financial profit, political reputation, business intelligence, and good working conditions.

Conflicts of Interests Different actors have different interests which can be in conflict each other. This section outlines the Conflicts of Interests between the actors of the proposed IT system architecture.

The individual interests of all actors is already described in the previous section and are not elaborated any further. The emphasis here is put on prominent existing conflicts, because they provide a foundation for vulnerabilities and subsequent threats.

System Complexity vs Privacy. System Providers offer a service to Local Authorities, so that Local Authorities can improve their public services. This task in it self has a high technical complexity and operates on privacy sensitive data provided by monitored Citizens.

Business Intelligence vs Privacy. Local authorities order a monitoring and mining system from system providers as much as possible, which allows them to produce business intelligence for public services. They are interested in the

successful usage of their data, although the interest of citizens lies in maintaining control over their data and protecting their rights to privacy.

Power of External vs. Privacy. Externals which are in a social relation to the citizen can have an interest in obtaining further information in order to gain power. In the most simplistic example this could be a man wanting monitor the activities of his spouse.

Financial Profit of External vs Privacy. Externals can gain financial profit from stealing privacy sensitive data. For example by selling raw contact information to advertisers or by selling mined data to insurance companies, or intermediaries like scoring companies. In such cases, citizens lose complete control over their data.

Financial Profit of External vs Reputation of System Providers. Externals have various business models as optional foundation for attacks on System Providers. For instance, they can try to invade the infrastructure for e-espionage reasons, to get control over servers to create a bot network or to steal user data. A successful attack proves the technical competence of system providers wrong and subsequently harms their professional reputation.

Political Activism vs Reputation of Local Authority. Besides monetary reasons, externals can be motivated by political reasons to attack the monitoring and mining system. Externals can break the system to make a political statement of their own, or they can steal user data to prove the system insecure. Both would harm the reputation of local authorities, who endangered the privacy of the citizens.

Vulnerabilities This section outlines the vulnerabilities (Figure 3) of the proposed monitoring and mining system. Note that vulnerabilities are not necessarily of technical nature. The weaknesses of IT systems are often created due to misuse or misconfiguration of the various components by one or more actors.

Insecure Infrastructure. The proposed monitoring system consists of many hardware and software components, each with its own concrete weaknesses. For instance, operating systems can be outdated or not subject to frequent updates or virus scans.

Insecure Data Transmission. The proposed monitoring and mining system uses HTTP to exchange data between the Sensor Collector, the Data Center and the Report Tool. Data can be intercept and read all sensitive information, which is send between the components e.g. passwords, raw sensor data and data mining results.

Unhappy Employees. An Employee that is frustrated with his situation for a long time period constitutes a security vulnerability. On the one hand he might want to harm his employer directly, on the other he is increasingly susceptible for social engineering.

Inadequate Access Rules. The proposed IT system infrastructure has various accesses to privacy sensitive data. System Provider staff has access to Data Center hardware and software like databases, web-servers and other inspection

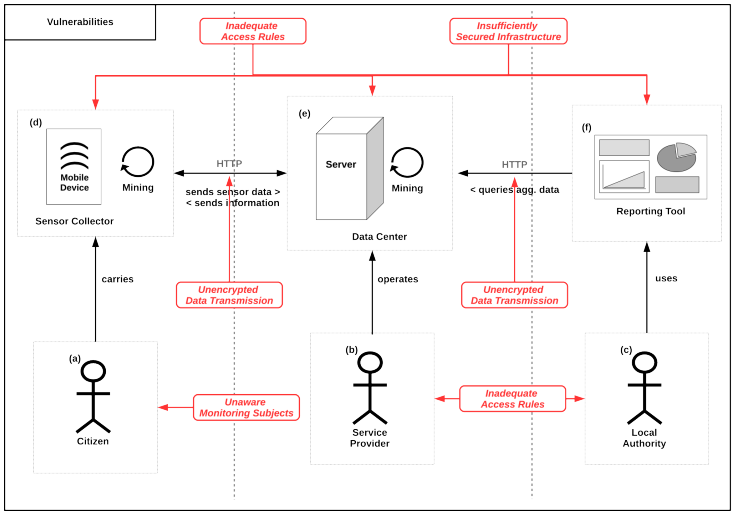


Fig. 3. Overview of vulnerabilities of the Live+Gov System; legend see 1

tools. Local Authority staff has access to the Report Tool. This all enables staff members to have potential access to privacy sensitive information.

Unaware Monitoring Subjects. We define privacy as one's ability to control information about oneself. In order to do that, monitored subjects need to know, that they are monitored, who monitors them, what information is recorded and for what purposes. Subjects who are not aware of these things cannot effectively preserve control and thus lose their privacy.

4.2 Step 2. Potential Analysis

The *Potential Analysis* displays the intended and unintended interactions of the components in the world view.

The intended interactions support the underlying business objectives. Unintended interactions can lead to threats. Obviously this provides a conflict of interest with the victim's business model, to keep the asset away from unauthorized access. From the point of view of the victim, an attack would be an unintended interaction.

The potential view consists of the following components:

- **Business Objectives.** Interaction of IT-system and actors that realize a business goals of the system owner were already described in Section 2.2 and 4.1.
- **Threats.** A threat is a potential interaction that destroys or harms assets of the system. Concrete realizations of threats can be *attacks* or *accidents*. Attacks are executed by an actor in response to a conflict of interest. Accidents are harmful interactions that are not willfully caused by an actor.

- **Chances/Risk.** Evaluation of chances and risks associated to the business objectives and threats. The risk associated to a threat is its expected loss.
- **Security Requirements.** Interactions (e.g. threats) that shall not occur within the system in order to achieve its business objectives. We included these security requirements into our analysis in Section 4.3.

Threat Specification A threat is a potential interaction of the components that targets an asset. We restrict ourselves to the case of attacks, and the asset of privacy. An attack is an interaction that is executed by an actor in response to a conflict of interest by exploiting a vulnerability of the system. The alternative interaction of accidents are less relevant for us, since the violation of privacy always requires an actor that takes advantage of personal data.

We describe threats for the citizen privacy in the Live+Gov system. This list can necessarily not be complete, but we make a best effort to cover the most relevant cases.

T1. Insufficient Control Features. As soon as collected data of Citizens is stored on System Provider servers, all control over that data is lost. Although, control capabilities for Citizens add to the system complexity, which could motivate to omit those. Therefore interests of Citizens and System Providers are in conflict, namely it is the Citizen's *Privacy vs. System Complexity* for System Providers. This is an abstract threat provoked by a general *Missing Privacy Awareness* in the minds of all actors in the Live+Gov context.

T2. Excessive Data Mining. The System Provider and/or the Local Authority secretly extract more private information from the collected data, than the Citizen agreed to. But results of the mining process create no disadvantages for Citizens, because there is no disclosure to third parties. This could be the case for a System Provider, who wants to test a new product and uses the pre-existing data collection. Or for a Local Authority, who wants to analyze the data collection regarding fare evasion. However, the Citizen has not agreed to such data processing nor could he, since it is conducted secretly. This disables a Citizen to control his data adequately. Thus there are two possible conflicts: *Privacy vs. Financial Profit of Serivce Provider* and *Privacy vs. Buisness Intelligence of Local Authority* The threat can be provoked by either lax data handling policies of both System Providers and Local Authorities, or a weak law enforcement of existing supervision. But the main issue, which can lead to such threats, is again a general *Missing Privacy Awareness*.

T3. Data Theft An External infiltrates infrastructure in order to steal personal data and sell it on the black market. Also the External might be motivated politically and wants to harm the reputation of the System Provider or the Local Authority. Such a successful attack could harm the reputation of both System Provider and Local Authority. Thus this threat is defined by three conflicts: *Privacy vs. Financial Profit*, *Reputation of Service Provider vs. Political Activism* and *Reputation of Local Authority vs. Political Activism*. Also this threat describes the classical scenario, where attacks are provoked by *Insecure Infrastructure* (SQL injection) and *Insecure Communication* (Packet Capture).

T4. Surveillance An External infiltrates infrastructure in order to obtain information about the citizen and exploit it directly. In this scenario the external is supposed to have some direct relationship to the citizen which motivates his interest to obtain personal information. Examples could be a public institution that wants to gain information about planned activities of the citizens (e.g. Nixon's Watergate scandal or the recent prosecution of Guardian journalists by GHCQ). In this threat the privacy interest of the citizen is in conflict with the aspirations for power over the citizen by the externals.

T5. Information Leak

Like an external person the Data Theft Scenario an employee of the service provider or the Local Authority has selfish interests to gain money, make political statements or harm his employer. In order to pursue this interest he can steal personal data and sell it or release it to the public. The corresponding conflicts of interests are: *Privacy vs. Financial Profit* of the Employee, *Reputation of Service Provider vs. Political Activism* of the Employee and *Reputation of Local Authority vs. Political Activism* of the Employee. The vulnerability constitutes of the existence of *Unhappy employees* itself and possibly *lax access rules* that enable the employee to obtain large amounts of data unnoticed.

T6. Social Engineering This scenario an external manipulates an employee of a Service Provider or the Local Authority to leak information to the external person. It is thus combination of the Data Theft and Information Leak scenario. The conflicts of interest are *Privacy vs. Financial Profit* of the External, *Reputation of Service Provider vs. Political Activism* of the external and *Reputation of Local Authority vs. Political Activism* of the external. The exploited vulnerabilities are, again, the existence of *Unhappy employees* and possibly *lax access rules* that enable the employee to obtain large amounts of data unnoticed.

Threat Risk Evaluation In this section we will associate to every identified threat a corresponding risk. A risk is the expected loss that is associated to the threat. Therefore, we have to quantify the likeliness of the threat to occure and the harm or loss done in this case. The quantification of likeliness will be solely based on rough judgment of the authors. The quantification of loss, will be made in a two step process. For each threat listed in the previous section, we have analyzed the affected personal data of the citizen. For each possible data type (e.g. GPS) we analyze the impact on the seven different types of privacy in Section 3.2. In combination we can quantify roughly the impact of each threat on the citizens privacy. Both evaluations are necessarily fraught with a high level of uncertainty.

For the quantification of the loss in case of a threat scenario we use the following rough calibration between 3 (high) and 0 (none). For the quantification of likeliness the following scale between 4 (always) and 0 (impossible) is used. The quantification of the risk, we add the values for loss and likeliness of the corresponding threats. Note, that loss and likeliness scales have a logarithmic character, so that addition of those scales corresponds to multiplication of the usual scales.

The likeliness, loss and the resulting risks assigned to the threats are discussed in the following paragraphs and summarized in Figure 4.

T1. Insufficient Control Features. The occurrence of this threat is dependent on the design on the system and given in our case, since we do not give the citizen control over his data once it is recorded. Therefore the Likeliness is evaluated as 4 (Always). The associated, risk is 1 Low on our scale, since no direct harm is done to the citizen by exploiting the data.

Hence the resulting risk is calculated as $4 + 1 = 5$.

T2. Excessive Data Mining. We assess the likeliness of excessive data mining to be 3 High, since these kind of analysis can be performed within the walls of the service provider, without somebody else noticing, and the service provider himself has an interest in this activity. The associated loss, on the other hand can be substantial (Medium 2).

Hence the resulting risk is calculated as $3 + 2 = 5$.

T3. Data Theft. The likeliness of a targeted attack by a third party is dependent on the popularity of the offered service and financial value of the captured information. Moreover, the amount of manual work required to infiltrate a custom build system is significantly higher that that of compromising a standard software solution. In the scenario we assume a moderate popularity in a single metropolitan area, with around 10.000 users and storage of data of only limited financial value (no addresses, no payment information). Therefore the likeliness assessment is $1 - 2$ (Low-Medium).

The harm of leaked information to a criminal party is 3 High. Hence the resulting risk is calculated as $4 - 5$.

T4. Surveillance. In the surveillance scenario an party related to the citizen, like a company where he is customer of, or a government agency, seeks to obtain sensitive information from our service.

The likeliness of such an intrusion is hard to asses, and depends again on the popularity of the service. If a high popularity is reached we have recently learned that spying by government agencies is very likely to occur. The barrier for companies that do not operate the infrastructure used to transmit the data a surveillance attack is however very hard to perform. Therefore we assess the likeliness of the threat with $1 - 2$ (Low-Medium).

The harm of leaked information to a related party is 3 High. Hence the resulting risk is calculated as $4 - 5$.

T5/6. Information Leak and Social Engineering.

In our scenario we assume that the culture and ethics inside the service provider company and local authority are very high, so that the information leak scenario has a likeliness of 1 (Low).

The harm of such an information leaked is 3 High, so that the resulting risk is calculated as 4.

Threat	Likeliness	Loss	Risk	Recommendation
T1. Insufficient Control Features	4	1	5	R1, R2
T2. Excessive Data Mining	3	2	5	R3, R4, R5
T3. Data Theft	1 - 2	3	4 - 5	R6
T4. Surveillance	1 - 2	3	4 - 5	R7
T5. Information Leak	1	3	4	R8
T6. Social Engineering	1	3	4	R8

Fig. 4. Live+Gov Risk Evaluation and Recommendations

4.3 Privacy Recommendations

In the preceding section we have identified the main risks for the users privacy. In this section we derive recommendations or requirements for a system that addresses these risks..

In order to address the threat with the highers risk, Insufficient Control (T1) of the citizen, we need to give the citizen back the control over its data inside the system. The most direct way to do this is to provide a web-based *Privacy Dashboard (R1)* which allows the citizen to view, edit and delete all information about his person that is stored inside the system. Also control applied processing and disclosure of the data to third parties should be given to the user, at least in the form of an opt-out or veto option.

A necessary pre-requirement for effective control of the citizen over his data is information and comprehension of the intended data capturing and processing steps. Therefore a *Privacy Policy (R2)* that is easily readable and contains all important information is essential and a legal requirement.

The threat with the second largest risk is (T2) Excessive Data Mining. Contrary to common belief, it is neither legal nor ethical to process personal data for by new methods or for new purposes that were not stated and explained to the citizen at the time of data collection. Also the common practice of obtaining far-reaching permissions from the citizens inside the privacy policy is neither an ethical or legal solution to the problem. To address this threat awareness about the limitations of data processors inside the company is a key element. As one mean to establish such a culture of privacy respect, we recommend to prepare an document called *Data Handling Guidelines (R3)* intended for internal use that explains the concrete processing steps and purposes that are permitted by the citizens. In particular the following information should be provided for each processing task: The name of the controller, purpose of processing, description of the data categories, recipients of the data if disclosed, transfer to third countries and a description of security of processing.

If further processing should be performed, it is necessary to seek additional permissions from the citizen. A simple email explaining the planned processing steps, and *asking for permission (R4)* would be enough for this purpose. The permission can be given via an embedded link that shall be followed in order to signal agreement.

An alternative measure to address the risk of excessive data mining is the *anonymization (R5)* of data. When all direct- or indirect links to the identity of the person are removed, no violation of the citizens privacy caused by arbitrary processing. However removing all such links is a challenging tasks, and full anonymity is often not achieved, cf. [10].

The protection from threat scenario (T3) Data Theft is a case of classical *IT infrastructure security (R6)*. The storage and processing infrastructure has to be secured using firewalls, up-to data software versions and proper authentication mechanisms.

The protection from threat scenario (T4) Surveillance focuses on the communication channels. They are target of wiretapping attacks by intermediaries or externals with access to the communication infrastructure. Strong *encryption (R7)* should be used to make it harder for externals to read the content of the transmitted data.

Threat scenarios (T5) Information Leak and (T6) Social Engineering target the vulnerability of unhappy employees. Therefore a trustful, *healthy company culture (R8)* should be maintained.

In summary we recommend the following measures to secure the citizens privacy:

R1 **Privacy Dashboard**. A tool which allows the citizen to view, edit and delete all data personal data that is stored in the system.
R2 **Privacy Policy**. A document, that informs the citizen about the collection and processing of personal information. It should at least contain the legally required information.
R3 Issue **Data Handling Guidelines** that explain the permitted processing methods and purposes.
R4 Ask the citizens for **permissions** before applying further processing via Email.
R5 **Anonymize** personal data before processing.
R6 **Securing of Storage** and Processing infrastructure using e.g. firewalls.
R7 **Securing communication** channels using encryption.
R8 Maintain a healthy, trustful **relationship** with your employees.

The mapping of these recommendations to the threats is summarized in Figure 4.

5 Conclusion

TODO: later.
* Key aspect control, can be given back to citizens via ‘‘Privacy Dashboard’’

References

1. Avizienis, A., Laprie, J.C., Randell, B., Landwehr, C.: Basic concepts and taxonomy of dependable and secure computing. Dependable and Secure Computing, IEEE Transactions on 1(1), 11–33 (Jan 2004)

2. Brandeis, L., Warren, S.: The right to privacy. Havard Law Review 4, 193–220 (1890)
3. BVerfG: decision about census, 15.12.1983, esp, c ii 1 a. and decision about online supervision by law enforcement, 27.2.2008.
4. Chambers, C.: Nsa and gchq: the flawed psychology of government mass surveillance (2013), http://www.theguardian.com/science/head-quarters/2013/aug/26/nsa-gchq-psychology-government-mass-surveillance [last access on 21.1.2015
5. Clarke, R.: Introduction to dataveillance and informatin privacy, and defintions of terms (1997), http://www.rogerclarke.com/DV/Intro.html, http://www.rogerclarke.com/DV/Intro.html [last access on 21.1.2015
6. EU: Directive 2002/58/EC of the European Parliament and of the Council Concerning the Processing of Personal Data and the Protection of Privacy in the Electronic Communications Sector (2002), http://europa.eu.int/eur-lex/pri/en/oj/dat/2002/l_201/l_20120020731en00370047.pdf [last access on 21.1.2015]
7. Finn, R.L., Wright, D., Friedewald, M.: Seven types of privacy. In: European Data Protection: Coming of Age. Springer Netherlands (2013), http://link.springer.com/chapter/10.1007%2F978-94-007-5170-5_1, [last access on 21.1.2015
8. Fried, C.: Privacy: A rational context. In: An Anatomy of Values, pp. 137–152. Havard Univ. Press (1970)
9. Grimm, R., Simić-Draws, D., Bräunlich, K., Kasten, A., Meletiadou, A.: Referenzmodell für ein vorgehen bei der it-sicherheitsanalyse. Informatik-Spektrum (2014), http://link.springer.com/article/10.1007%2Fs00287-014-0807-3, [last access on 21.1.2015
10. Krumm, J.: A survey of computational location privacy. Personal and Ubiquitous Computing 13(6), 391–399 (2009)