Reality Sensing, Mining and Augmentation
for Mobile CitizenGovernment Dialogue
FP7-288815

D1.3 - Privacy Aware Sensor Data Storage and Miner

Dissemination level:	PU - Public
Contractual date of delivery:	Month 30, October 2014
Actual date of delivery:	Month 30, October 2014
Workpackage:	WP1 - Reality Sensing and Mining
Task:	T1.3, T1.4
Type:	Prototype
Approval Status:	PMB Final Draft
Version:	12
Number of pages:	39
Filename:	D1-3.tex

Abstract

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

This work was supported by the EU 7th Framework Programme under grant number IST-FP7-288815 in project Live+Gov (www.liveandgov.eu)

Copyright 2013 Live+Gov Consortium consisting of:

- Universitt Koblenz-Landau
- Centre for Research and Technology Hellas
- Yucat BV
- Mattersoft OY
- Fundacion BiscayTIK
- EuroSoc GmbH

This document may not be copied, reproduced, or modified in whole or in part for any purpose without written permission from the Live+Gov Consortium. In addition to such written permission to copy, reproduce, or modify this document in whole or part, an acknowledgement of the authors of the document and all applicable portions of the copyright notice must be clearly referenced.

All rights reserved.

History

Version	Date	Reason	Revised by
01	2014-07-17	Outline	Heinrich Hartmann

Author list

Organization	Name	Contact Information
UKob	Heinrich Hartmann	Phone: +49 261 287 2759 Fax: +49 261 287 100 2759 E-mail: hartmann@uni-koblenz.de
UKob	Christoph Schaefer	Phone: +49 261 287 2786 Fax: +49 261 287 100 2786 E-mail: chrisschaefer@uni-koblenz.de

Executive Summary

The deliverable is accompanied with source code of the components in Java, API documentation (javadoc), and pre-compiled packages for direct installation and testing on mobile devices.

Abbreviations and Acronyms

AIDL	Android Interface Description Language
AJAX	Asynchronous JavaScript and XML
ALC	Attribute Language with Complement
API	Application Programming Interface
GeSA	Geographical Semantic Analysis
GPS	Global Positioning System
GSM	Global System for Mobile Communications
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
ID	Identifier
IEEE 802.11	see WIFI, WLAN
JSON	JavaScript Object Notation
LAN	Local Area Network
REST	Representational State Transfer
RF-ID	Radio-Frequency Identification
SDCF	Sensor Data Collection Framework
SQL	Structured Query Language
SVM	Support Vector Machine
TBox	Terminological Box
UI	User Interface
URL	Uniform Resource Locator
UUID	Universal Unique Device Identifier
WP	Work Package
WIFI	Wireless Fidelity (IEEE 802.11), WLAN
WLAN	Wireless Local Area Network
XML	Extensible Markup Language

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Privacy

The goal of this document to analyze and identify the threads 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 the last part we describe how these measures are implemented in the Live+Gov toolkit.

As privacy is a very general and hard to grasp term, we need to fix a definition of privacy that is suitable for our needs. As background information we include an overview about historical treatments of privacy as well as legal regulation of privacy in the European Union. Based on this we propose a definition of privacy as *control over personal data*, and introduce a taxonomy of privacy attributes that give specify the term *personal data* in the context of mobile sensor data collection.

The complexity of our systems and the variety of threads make a great number of counter measures plausible. We approach this complexity with the aid of a general security analysis model developed in [Grimm]. We give a brief introduction to this model and perform a IT Security Analysis with respect to the privacy asset for our system.

1.1.1 Histrocial Approaches to Privacy

1.1.1.1 Aristotle

Public Sphere vs. Private Sphere of politics. This distinction is concerned with the politic (public) life of politicians in early democracies opposed by their domestic (private) life.

1.1.1.2 Warren and Brandeis

The Right to Privacy, 1890: Establish the term Informational Privacy. They argue innovations like photography and a growing press create the need for a new right to privacy, ”the right to be left alone”.

The term informational privacy as conceptual foundation of privacy originates from the juristic discussion The Right to Privacy (1890) by Warren and Brandeis. The arise of new technologies like photography combined with a growing press created an increased publicity. Warren and Brandeis were worried that existing law like castle doctrines or libel and slander could not suffice. They felt a need for a new right to ”protect the extent to wich one’s thoughts, sentiments, and emotions could be shared with others” [1]. This right would be ”the right to be left alone”.

The right to ”protect the extent to wich one’s thoughts, sentiments, and emotions could be shared with others” by Warren and Brandeis was later extend to cover arbitrary information and is now kown as the privacy concept ”... the control we have over information about ourselves ...” [2].

1.1.1.3 Fried

C. Freid, An Anatomy of Values, Harvard Univ. Press, 1970

Defines privacy as the ability to control information about oneself. He argues that limited knowledge does not automatically create privacy.

1.1.2 Legal Aspects of Privacy

* Present legal view on privacy from European angle: "Data Protection Directive" and Implementation.

* Some US Law

1.1.2.1 European Convention on Human Rights - Article 8

"(1) Everyone has the right to respect for his private and family life, his home and his correspondence. (2) There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others." [3]

"The European Court of Human has given this article a very broad interpretation in its juresprudence." [2]

1.1.2.2 EU Data Protection Direvtive

Directive 95/46/EC

EU directives in general do not hold any direct legal binding for EU citzens. Member states have to create their own legal implementation.

"The directive regulates the processing of personal data regardless of whether such processing is automated or not." [1] where:

- **personal data** is "any information relating to an identified or identifiable natural person [...] (art. 2a)" [1]
- **processing** is "any operation or set of operations which is performed upon personal data [...] (art. 2b)" [1]

The "identified or identifiable natural person" is called **data subject**.

The sole responsibility for compliance is held by so called **controllers**. A controller is an actor "which alone or jointly with others dertimines the purposes and means of the processing of personal data; (art 2d)" [1]

(**Note:** this rule is applicable "whenever the controller uses equipment situated within the EU"[1]. This will hold for every e-commerce provider **inside or outside** of the EU because the customer's computer is *situated within the EU* anyways.)

"Personal data should not be processed at all, except when certain conditions are met" [1]. These conditions incorporate the seven OECD princepals and are futher categorized in the categories:

- **Transparency**: "The data subject has the right to be informed when his personal data is being processed [...] (art. 10 and 11)" [1] and he has the right how and by whom it is processed. Additionally other speficied requirements have to be met, i.e. giving consent.
- **Legitimate prupose**: "Personal data can only be processed for specified explicit and legitimate pruposes and may not be processes futher in a way incompatible with those pruposes (art. 6b)" [1].
- **Proportionality**: "Personal data may be processed only insofar as it is adequate [...] in relation to [its] pruposes [...]; every reasonable step must be taken to ensure that data wich are inaccurate or incomplete [...] are erased or rectified (art. 6)" [1]

(**Note:** The principles incorporated by the EU directive seem to closely map to *the seven principles for user privacy control*.)

Member states have to create independent **supervisory authorities** for personal data processing where controllers have to register before they start to process any data. The record has to be stored in public register [1].

Data transfer to thrid countries outside of the EU requires those countries to have a similar data protection level. Exceptions exists:

- Save Harbor
- Passenger Name Record Aggreement

all between the US and the EU, although the US has no comparable law of data protection. The US Supreme Court regards privacy as **implicit** right granted with the First Amendment. However, european law states privacy as **explicit** right (ECHR Art. 8). Additionally the US government follows a doctrine which favors the economy to implement self-regulation [1].

1.1.2.3 Implementation of the Data Protection Direvtive

1.1.2.3.1 Germany

Germany implements Directive 95/46 with the *Federal Data Protection Act* from 2001 known as *Bundesdatenschutzgesetz (BDSG)* [1][2][6]. However, Germany has violated the directive in two points:

1. The BDSG has become effective three years too late, thus the EC filed a treaty violation proceeding against Germany [6].
2. The BDSG does not implement **independet** supervisory authorities. The *Bundesdatenschutzbeauftragter* is subordinate to the Ministiry of Interior. Although he is not subject to technical oversight (*Fachaufsicht*), he is subject to staff supervision by the government (*Rechtsaufsicht* durch die Bundesregierung und *Dienstaufsicht* durch das Innenministerium) and budget oversight (including approval of employees) by the ministry [7]. Thus the EC filed a treaty violation proceeding against Germany, again [6]. In March 2010 Germany was found guilty of violation of Directive 95/46 by the ECJ [6].

(**Note:** Germany still violates point 2!)

The states of Germany have their own implementation of Directive 95/46 (*Landesdatenschutzgesetze*). Federal public authorites are only bound to their federal law. [2]

Churches are not subject the BDSG. [2]

- **Roman Catholic Church:** Anordnung ber den kirchlichen Datenschutz (KDO)
- **German Protestant Churches (Synode der Evangelischen Kirche in Deutschland):** Kirchengesetz ber den Datenschutz der Evangelischen Kirche in Deutschland

Data subjects have the right to...

- ... **disclosure** of
- wether data about them is stored and processed
- which data is stored and processed
- the data sources
- ... **correction** of false data
- ... **file complaints** to the supervisory authority
- ... have data **deleted** or **blocked**, but controllers can prohibit deletion in favour of blocking
- ... **Decline** third party access to the data

1.1.2.3.2 United Kingdom

The UK implements Directive 95/46 with the *Data Protection Act 1998 (DPA)* [1][3].

The act is knwon for its high complexity: a manual record of phone numbers for business purposes could be hold subject to the DPA [3]. Although the act seems to fully cover the directive. Even higher restriction apply for *"sensitive personal data"* (race, ethnicity, politics, religion, trade union status, health, sex life or criminal record), i.e. **consent** must be given freely and has to be explicit. [3]

"The Act's definition of"personal data" covers any data that can be used to identify a living individual. Anonymised or aggregated data is not regulated by the Act, providing the anonymisation or aggregation has not been done in a reversible way. Individuals can be identified by various means including their name and address, telephone number or Email address. The Act applies only to data which is held, or intended to be held, on computers ('equipment operating automatically in response to instructions given for that purpose'), or held in a 'relevant filing system'.

In some cases even a paper address book can be classified as a 'relevant filing system', for example diaries used to support commercial activities such as a salesperson's diary.

The Freedom of Information Act 2000 modified the act for public bodies and authorities, and the Durant case modified the interpretation of the act by providing case law and precedent.

The Data Protection Act creates rights for those who have their data stored, and responsibilities for those who store, process or transmit such data. The person who has their data processed has the right to:

- *View the data an organisation holds on them. A 'subject access request' can be obtained for a nominal fee. As of January 2014, the maximum fee is 2 for requests to credit reference agencies, 50 for health and educational request, and 10 per individual otherwise,*
- *Request that incorrect information be corrected. If the company ignores the request, a court can order the data to be corrected or destroyed, and in some cases compensation can be awarded.*

- *Require that data is not used in any way that may potentially cause damage or distress.*
- *Require that their data is not used for direct marketing.“* [3]

1.1.2.3.3 France

France implements Directive 95/46 with Law 2004-801 modifying law 78-17 of 6.1.1978 (*Loi n 2004-801 du 6 aot 2004* modifying *loi n78-17 relative l'informatique, aux fichiers et aux liberts*) [1][10][11].

1.1.2.3.4 United States of America

The USA do not implement the directive, nor is there any obligation for them to do so. However, companies subject to US jurisdiction can be certified to comply with the seven principles enforced by Directive 95/46 (the seven OECD recommendations) [5]. Thus, those companies will act as *safe harbors*. Without certification foreign companies are not allowed to send customer data back home [9].

(**Note:** *Safe harbor* in general is the legal concept to regulate that a certain conduct will be deemed [4], but in germany the term is commonly used as synonym for the agreement between the EU and the US regarding Directive 95/46.)

1.1.3 Privacy Definition and Taxonomy

1.1.3.1 Defining Privacy

Defining privacy is a challenge which seems impossible. This is well put to words by Serge Gutwirth, who notes: *“The notion of privacy remains out of the grasp of every academic chasing it. Even when it is cornered by such additional modifiers as ‘our’ privacy, it still finds a way to remain elusive.”* [2].

Many privacy researchers seem to “forcus on the ways in which privacy can be infrenged” [1]. Thus they try to create taxonomies of *privacy harms* instead of taxonomies of *privacy types*. Those two differ in the respect that the former focuses on threats to prohibit whereas the latter focuses on values to protect. So one should rather evaluate what aspects are precious about privacy and develop measures to ensure their security than only forbid single actions against privacy [1].

We follow Fried in defining privacy as the ability to control data about personal information.

* What is the relation to legal aspects?

* What should be personal information?

* Use Seven Privacy Types be Friedewald, Finn and Wrigtht.

1.1.3.1.1 The Seven Types of Privacy

Seven Types of Privacy based on four types of privacy by Clarke. Clarke's four types are outdated by contemporary technologies and no longer adequate. In order to fix this Friedewald,

Finn and Wright extend the former four to the now introduced seven types privacy as follows:

1. **Privacy of the Person.**

   This type is the right keep body functions and body characteristics private [1]. It maps the one by Clarke.

   - Body Characteristics
   - Weight
   - Height
   - Other body measures (shoulder width)
   - Fingerprints
   - DNA Sequence
   - Medical Conitions
   - Orthopedic conditions (e.g. limping)
   - Having a cold
   - ...

1.1.3.1.1.1 Privacy of Behaviour and Action This type is also concerned with the "protection against disclosure of personal matters" [1] through behaviour, however Clarke's distinction between "casual observation [...] systematic recording and storage of information about those activities" [1]is lifted.

- regular visit at church, backery, doctors
- sexual habits
- political activities

1.1.3.1.1.2 Privacy of Communication It "aims to avoid the interception of communications" [1] either electronic or face-to-face.

- Briefgeheimnis (in german Legislation)
- Email contents
- Personal direct communication
- Right to free discussion, i.e. without third parties listening

1.1.3.1.1.3 Privacy of Data and Image This type is concerned with "making sure that individuals's data is not automatically available to other individuals and organisations" [1]. This is Informational Privacy in an intuitive sense.

- Telefonnummer
- IP Addresse
- Public-administrative Data (Date of Birth, Melderegister)
- Data held by organizations, like Banks or Insurance Companies
- All data that is stored in online services (facebook)

1.1.3.1.1.4 Privacy of Thoughts and Feelings This type is the right "not to share their thoughts or feelings or to have those thoughts or feeling (sic!) revealed" [1]. It is concerned

with (automatic) emotion detection. This type is the counterpart the *Privacy of the Person* like body and mind are counterparts of one another (dualism).

- Current feeling: Depression, tiredness, stressed, awake
- Thoughts in general

1.1.3.1.1.5 6 Privacy of Location and Space This type is the right "to move about in public or semi-public space withoug being identified, tracked or monitored" [1]. Additionally this type is concerned with the protection of one's home and private places ("right to solitude" [1]).

- GPS position tracking
- Location of home address

1.1.3.1.1.6 Privacy of Association This type is the right "to associate with whomever [one] wish, withoug being monitored" [1]. It is concerned with the protection against the automatic record of one's contacts. It does not imply that one is monitored because of the associations.

- Friends
- Joining of orgnaizations (e.g. political parties)

1.2 IT Security

1.2.1 Basic Terminology

1.2.1.1 threat

threat is a class of potential events whose manifestation can cause damage or harm.

1.2.1.2 vulnerability

A weakness which can cause the loss of security.

1.2.1.3 hazard

Concrete risk due to one or more concrete vulnerabilities **AND** corresponding threats.

```
threat + vulnerability -> hazard
```

1.2.1.4 incident

An event caused by a threat manifestation against one or more corresponding vulnerabilities.

```
threat manifestation + vulnerability -> hazard
```

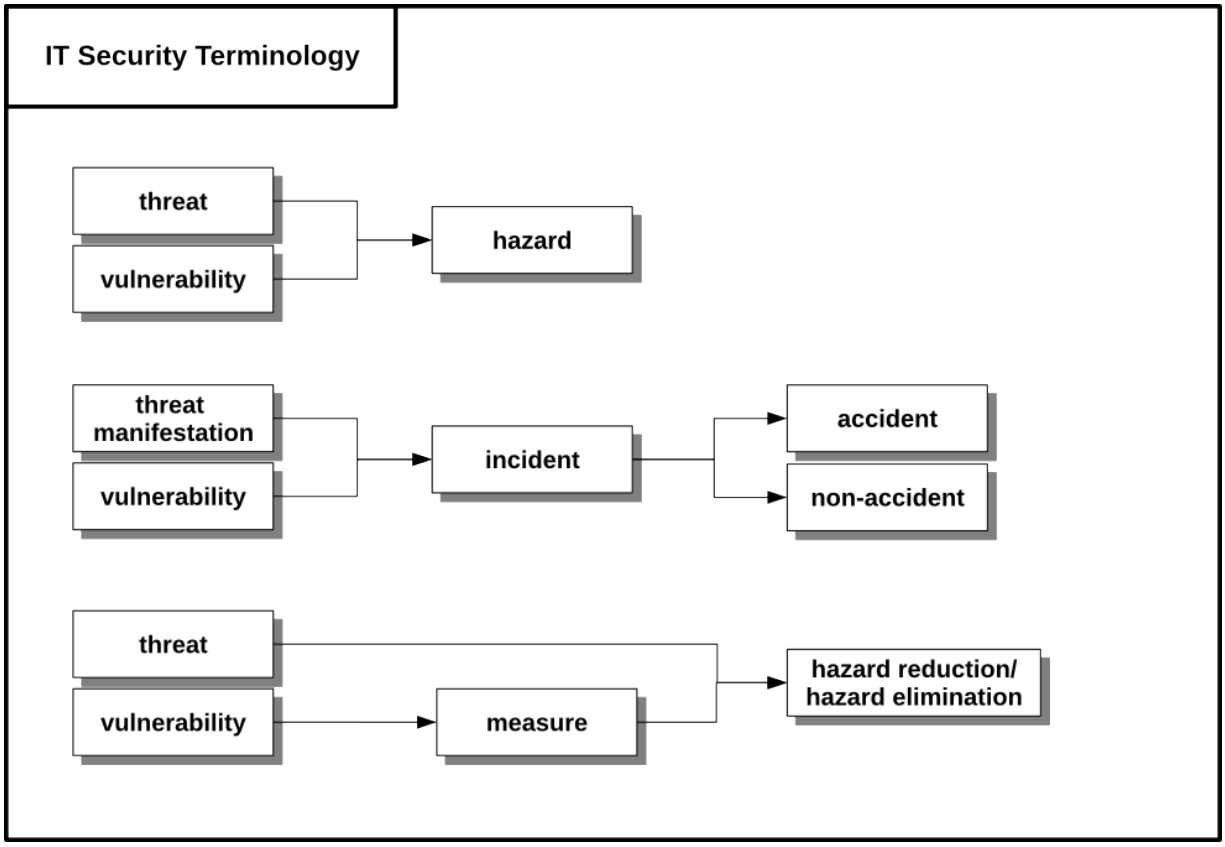


Figure 1: Basic IT Security Terminology

1.2.1.5 accident

An incident which did cause damage or harm. (Translates to *Schandensfall/Zwischenfall/Strfall* in german)

1.2.1.6 non-accident

An incident which **did not** cause any damage or harm.

1.2.1.7 measure

A well defined set of one or more activities which reduce or eliminate vulnerabilities.

1.2.2 IT Security Analysis according to Grimm et. al

Grimm et al. create a reference model for conducting IT security analyses consisting of:

- an **ontology**, which aims to organize common security terminology in a reasonable and practical way
- and a systematic analysis **procedure** based on that ontology

1.2.2.1 Ontology

1.2.2.1.1 Views

The ontology is organized in four views:

- **World:** Contains all components describing the current state.
- **Potential:** Contains all components describing both the desired and dreaded state.
- **Plan:** Contains all conceptual components required to develop sufficient measures to achieve and secure the desired state **or** to create the dreaded state. The **Plan** view *realizes* the **Potential** view of the system. (**Note:** this should rather be translated with *Concept*)
- **Event:** Contains all actual operations and events through out the production phase. The **Event** view *instantiates* the **Plan** view of the system.

1.2.2.1.2 Components

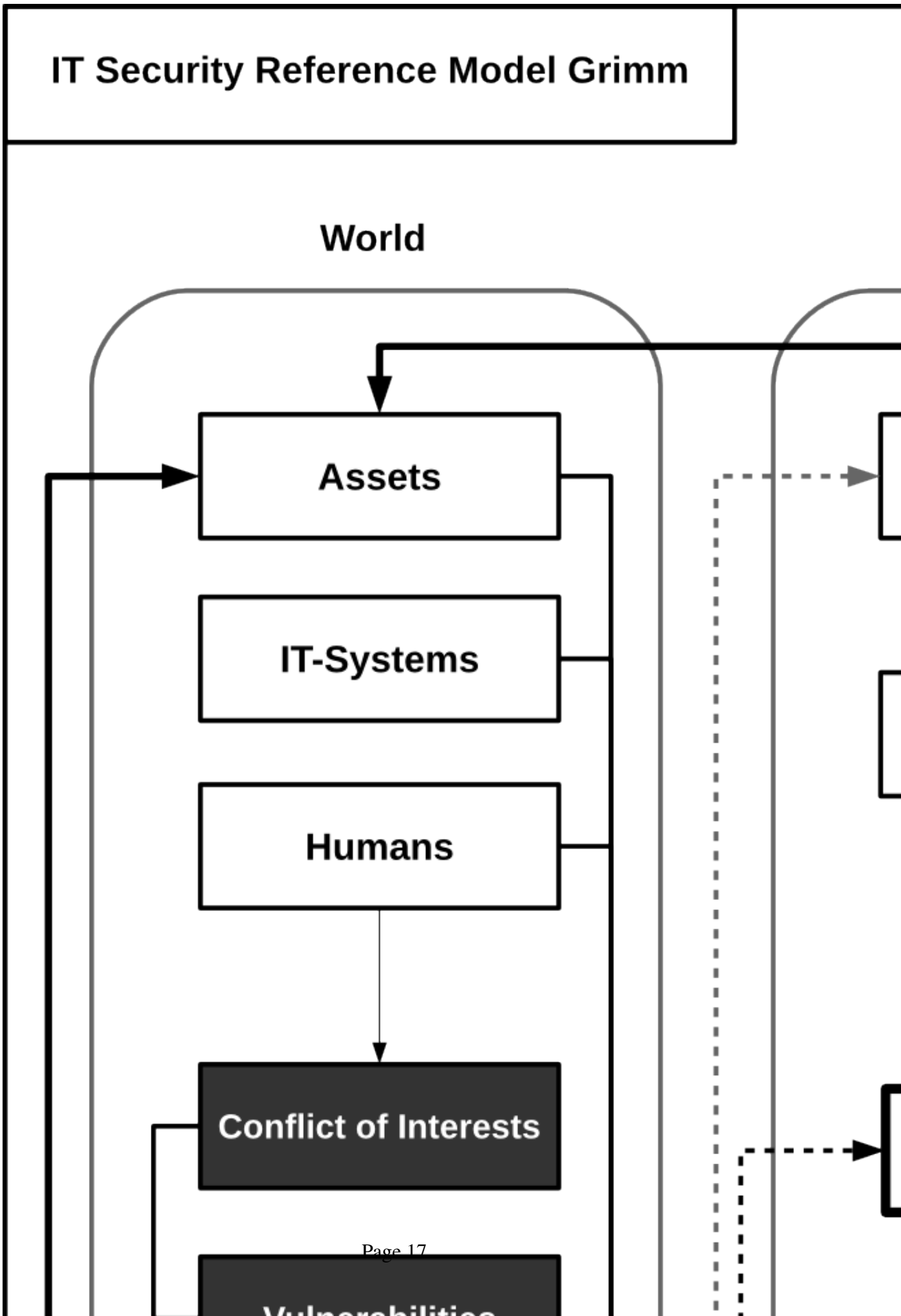
1.2.2.1.2.1 World

- **Assets:** Things of value to one or more stakeholders. The value can be *"hard"* (money, data, etc.) or *"soft"* (trust,privacy,etc.).
- **IT-Systems:** The IT-Systems under study.
- **Humans:** All identifiable actors of the system under study. Many problems arise due to misunderstandings during man-machine interaction.
- **Conflicts of Interests:** Different actors have different interests. Those interest can be in conflict. A trival conflict is the *Attacker-Attackee*-Conflict: a service provider offers private data storage, therefore the provider is interested in having the access restricted. An attacker is naturally interested in easy access in order captialize the stolen data. However, heavy security restriction are also a burden for maintenance staff, as they are interested in having an easy life. This kind of conflict can create vulnerabilities.
- **Vulnerabilities:** All indentifiable weaknesses in the current system layout.
- **Interactions:** Assets, IT-Systems, Humans and Vulnerabilites are in continous interaction with each other in order to *support* the **Business Objectives**. Those interactionscan also *lead to* **Threats**, i.e. having unencrypted communication with a server. Therefore all interactions which occur in the system's outline have to be documented.

1.2.2.1.2.2 Potential

- **Chances/Risks:** Chances and Risks *affect* **Business Objectives** (and **Threats**).
- **Business Objectives:** Chances the system under study. This describes the desired state.
- **Threats:** Risks for business objectives, this can simply be the integrety of the IT-System. Threats *target* one or more **Assets** and violate **Security Requirements**. This describes the dreaded state.
- **Security Requirements:** A set of distinct requirements for a safe and secure system. Security requirements *hold for* one or more **Assets**. This describes the potential state after this procedure.

1.2.2.1.2.3 Plan


IT Security Reference Model Grimm
World
Assets
IT-Systems
Humans
Conflict of Interests

- **Business Model:** The plan to achieve **Business Objectives**
- **Accident Scenario:** This is either the plan of a certain attack or the concrete outline of random disaster. Although the latter is most likely hard to anticipate.
- **Attack Technique:** A specific technique or technology to attack IT-Systems (Man in the Middel, Phishing, etc.). (**Note:** This component is not conform to the glossary as the term *accident* there covers both disaster and attack!)
- **Security Measures:** The plan to achieve a safe and secure system. Each **Security Mesaure** *opposes* an **Accident Scenario** or an **Attack Technique**.

1.2.2.1.2.4 Event

- **Business Process:** The actual, running instance of the **Busieness Modell**
- **Accidents:** All actually happened accidents.
- **Attacks:** All actually happened attacks. (**Note:** This component is not conform to the glossary as the term *accident* there covers both disaster and attack!)
- **Security Operations:** Instances of **Security Measures**. Security Measures *secure* the system against **Accidents** and **Attacks**.

1.2.2.2 Procedure

The analysis procedure is an incremental and iterative process following the four views of the previously described ontology.

1.2.2.2.1 Step 1. World Analysis

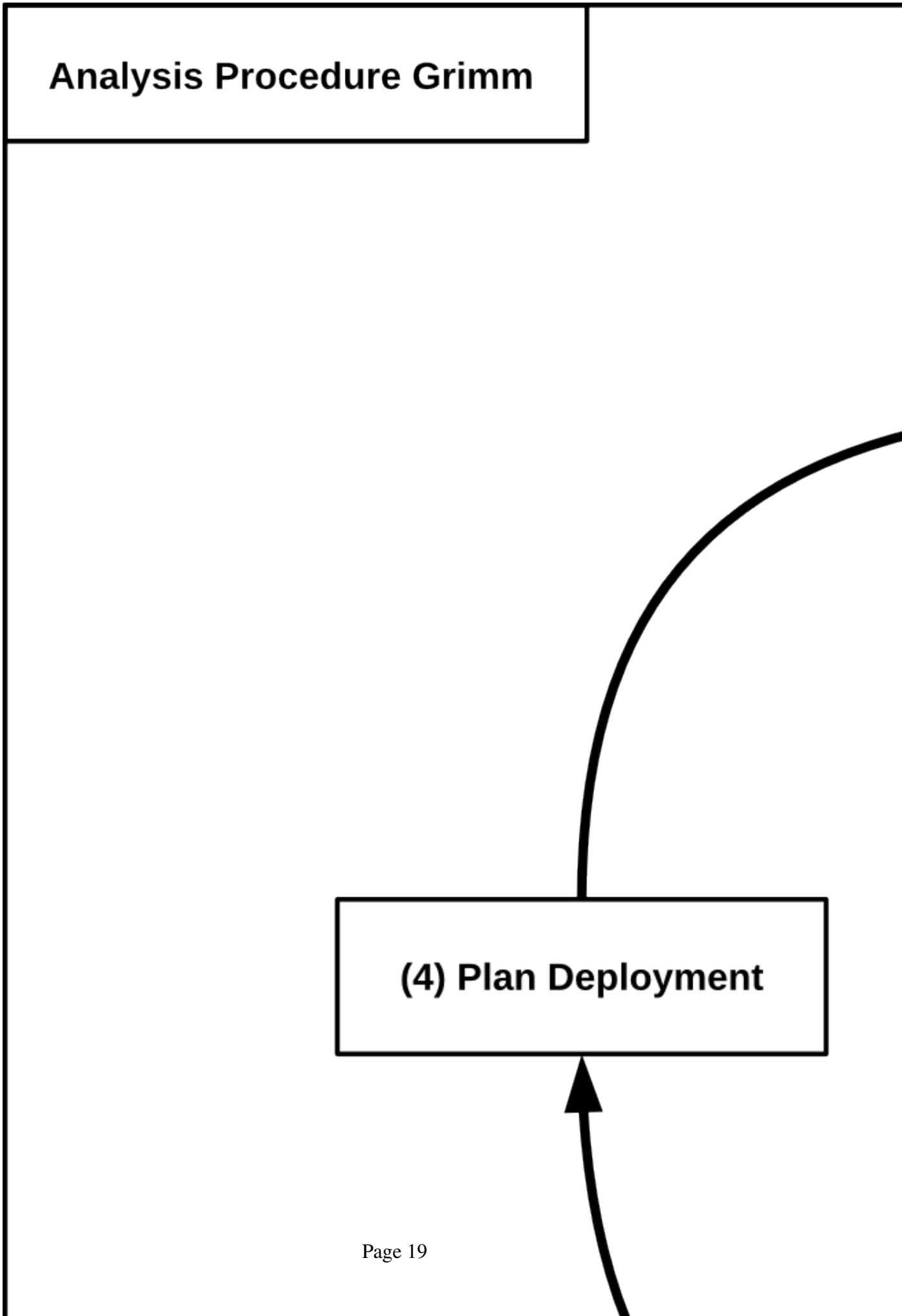
At first, one has to outline the current state of the system under study. This includes description of:

- all **Assets** which must be protected
- all relevant **IT-Systems**
- all involved **Humans** and their **Conflicts of Interests**
- all known **Vulnerabilities**
- and all important **Interactions** between the former components.

1.2.2.2.2 Step 2. Potential Analysis

Secondly, one needs to outline the potential state of the system under study. This includes both the dreaded (**Threats** and **Risks**) and the desired (**Business Objectives**, **Chances**, **Security Requirements** state. This step produces four artefacts:

- a *threat specifiation*, which identifies the **Threat** and its targeted **Assets**
- a *threat risk evaluation*, which detrmines the likelyhood of a threat manifestation
- a *hazard matrix*, which maps threats to known **Vulnerabilities** and identifies potential hazards
- a *security requirement specifiation*, which specifies requirments in order to deal with identified hazards


Analysis Procedure Grimm
(4) Plan Deployment

1.2.2.2.3 Step 3. Plan Development

Based on **Step 2.**, the identified hazards are used alongside realistic **Accident Scenarios** (**Attack Techniques**) to create a *risk matrix*. With this matrix it is possible to decide if the risk is acceptable or not. Together with the previously specified **Security Requirements**, the matrix is used to define adequate **Security Measures**. Like a **Business Model** is an abstract concept to achieve **Business Objectives**, this step creates an abstract concept to improve the system's security.

1.2.2.2.4 Step 4. Plan Deployment

Finally, the **Security Measures** have to be implemented. Additionally, all **Business Operations**, **Accidents** (**Attacks**) and executed **Security Operations** will be recorded through out the production phase.

The implementation of **Security Measures** evantually changes the **World** and renders the conducted analysis outdated. So this analysis procedure needs to be conducted again.

1.2.2.3 Abstraction Levels of the Reference Model

The Reference Model can be used on different levels of abstraction. This means each component can be used within a wide range of granularity, for instance the security measure *Encryption* can be explored in genral or on the level of different concrete encryption tools; or on the even finer level of concrete algorithms.

The utilized abstraction level is not important for the analysis procedure, it depends on the intended audience for the analysis. However, it is important to use one abstraction level consistently through out the analysis.

1.3 Live+Gov Privacy Protection Analysis

1.3.1 Step 1. World Analysis

1.3.1.1 Assets: Privacy

We focus on the Asset privacy as described in Section **??**

1.3.1.2 IT-Systems & Humans

1.3.1.2.1 Server Side Mining Scenario (Figure 4)

A citizen (a) carries a mobile device (b) running the L+G client. The mobile device collects sensor data and sends it to the L+G Data Cetner (c). The L+G Data Center also sends beneficial information for citizens to the mobile device.

The L+G Maintenance Crew (d) maintaince the L+G Data Center.

The Aggregation Monitor (e) queries the L+G Data Center for aggregated data. Local Authorities (f) use the Aggregation Monitor to get information in order to improve public services. **In this scenario the mobile the server conducts data mining on aggregated user data and additionally trys to combine it with publicly accessible data to enhance the results.**

When no connection to the L+G servers is available, the mining end-products are stored on the mobile device.

1.3.1.2.2 Mobile Mining Scenario (Figure 5)

A citizen (a) carries a mobile device (b) running the L+G client. The mobile device collects sensor data and sends it to the L+G Data Cetner (c). The L+G Data Center also sends beneficial information for citizens to the mobile device. **In this scenario the mobile device also conducts data mining previously to sending the results to the L+G Data Center.**

The L+G Maintenance Crew (d) maintaince the L+G Data Center.

The Aggregation Monitor (e) queries the L+G Data Center for aggregated data. Local Authorities (f) use the Aggregation Monitor to get information in order to improve public services.

When no connection to the L+G servers is available, the mining end-products are stored on the mobile device.

1.3.1.3 Conflict of Interests (Figure 6)

Citizen vs L+G Maintenance Crew

- citizens want to be unidentifiable
- citizens want certain informations about them to be secret
- technical staff needs more or less unrestricted access to data in order to maintain the system

Citizen vs Local Authority

- citizens want to be unidentifiable
- citizens want certain informations about them to be secret

L+G Maintenance Crew vs Local Authority

- staff wants proper payment
- local authorities want to reduce cost

Criminal vs Citizen

- citizens want to be unidentifiable
- citizens want certain informations about them to be secret
- criminals want to posses certain information about a citizen

Criminal vs L+G Maintenance Crew

- criminals want to harm the system
- technical staff has to keep the system safe and sound

Criminal vs Local Authority

- criminals want to harm the system
- criminals want access to collected data
- local authorities have to keep collected data restricted
- local authorities want a stable system

1.3.1.4 Vulnerabilities

1.3.1.4.1 Combined Scenario (Figure 7)

A citizen (a) carries a mobile device (b) running the L+G client. The mobile device collects sensor data and sends it to the L+G Data Cetner (c). The L+G Data Center also sends beneficial information for citizens to the mobile device.

The L+G Maintenance Crew (d) maintaince the L+G Data Center.

The Aggregation Monitor (e) queries the L+G Data Center for aggregated data. Local Authorities (f) use the Aggregation Monitor to get information in order to improve public services.

A criminal (g) threatens the L+G system by corrupting the L+G Maintenance Crew or Local Authority, or by forcing access to L+G system's hardware, software or communication. (It is also possible for criminals to corrupt the citizen, although technically this would be no threat to privacy but rather an exercise of privacy [1].)

1.3.1.4.2 List of Vulnerabilities

The vulnerabilities of Live+Gov system outline are:

- unencrypted data transmission (Man In The Middel)
- insecure mobile devices (Exploit)
- insecure servers (Exploit)
- inadequate access rules for
  - mobile devices (citizens)
  - servers (staff)
  - applications (local authority)
- corrupt/unhappy authorities or staff (**Could be actually a threat not a vulnerability**)

(**Note:** The vulnerabilities seem to mostly open possibilities to violate the **Privacy of Data and Image**, as this type is concerd with making data *automatically* available to others.)

1.3.2 Step 2. Potential Analysis

1.3.2.1 Threats

- **Unauthorized access to privacy sensitive data**, caused by
  - **Excessive Data Mining** Linking sensor data provided by citizens with additional sources can produce more privacy sensitive data.
  - **Corrupt Local Authorities** Local Authorities have certain data access, they could hand over this access for monetary reasons.
  - **Corrupt/Unhappy Staff** Staff members also have a certain data access and they also could hand over this access for monetary reasons or as a form of payback for unfair treatment.
  - **Competent Attackers** Competent Attackers are always a threat for IT-systems. Normaly they have monetary reasons to exploit a system, but there is also a possibility for proof-of-concept like attacks.

1.3.2.1.1 Implicit Privacy Type Violation (Figure 8)

1.3.2.1.1.1 Privacy of The Person The Privacy of The Person is concerned with one's biometric privacy. If this type is violated, following implicit violations are possible:

- **Privacy of The Person:** reflexive violation
- **Privacy of Behaviour and Action:** none
- **Privacy of Communication:** none
- **Privacy of Data and Image:** none
- **Privacy of Thoughts and Feelings:** Some psychological diseases (e.g. depression) have physiological impact. Such physiological patterns could be detected.
- **Privacy of Location and Space:** none
- **Privacy of Association:** none

1.3.2.1.1.2 Privacy of Behaviour and Action The Privacy of Behaviour and Action is concerned with one's privacy regarding social activities (religious, political, sexual, ...). If this type is violated, following implicit violations are possible:

- **Privacy of The Person:** Religious practices may include body modifications (e.g. circumcision).
- **Privacy of Behaviour and Action:** reflexive violation
- **Privacy of Communication:** none
- **Privacy of Data and Image:** none
- **Privacy of Thoughts and Feelings:** Social activities in general depend on a certain intellectual attitude. Such an actvity is the expressions of such an attitude.
- **Privacy of Location and Space:** none
- **Privacy of Association:** Recording religious, politcal or sexual activities can reveal association with churches, political parties or sexual partners.

1.3.2.1.1.3 Privacy of Communication The Privacy of Communication is concerned with not havin such communication (corresponence or vis-a-vis) intercepted. This is very broad type of

privacy. Depending on the contens of the intercepted communication every other type can be violated:

- **Privacy of The Person:** Communication about body characteristics.
- **Privacy of Behaviour and Action:** Communicatio about social activities.
- **Privacy of Communication:** reflexive violation
- **Privacy of Data and Image:** Communication containing one's passwords or other sensitive data.
- **Privacy of Thoughts and Feelings:** Communication of thoughts and feelings, e.g. wiretapping a flirt or a catholic confession ritual.
- **Privacy of Location and Space:** Interception of face-to-face communication is only possible if one's location and space is violated (wiretapping).
- **Privacy of Association:** Communication about one's associations (family members, churches, etc.).

1.3.2.1.1.4 Privacy of Data and Image The Privacy of Data and Image is concered with one's data not beeing automatically availabel to others. This also is a very broad type of privacy. Depending of the data or image contents erery other type can be violated:

- **Privacy of The Person:** Images or stored biometric information reveal one's physical characteristics.
- **Privacy of Behaviour and Action:** Images or diaries can reveal one's social activities.
- **Privacy of Communication:** Modern communication systems usually contain some sort of archive function, e.g. E-mail clients do not automatically delete messages. Such messages are data and reveal one's communication.
- **Privacy of Data and Image:** reflexive violation
- **Privacy of Thoughts and Feelings:** Images can show one's emotional state.
- **Privacy of Location and Space:** Images can reveal one's location, e.g. making a picture in front of the Eifel Tower.
- **Privacy of Association:** E-mail data can also reveal association.

1.3.2.1.1.5 Privacy of Thoughts and Feelings The Privacy of Thoughts and Feelings is concerned with keeping such thoughts and feelings secret. If this type is violated, following implicit violations are possible:

- **Privacy of The Person:** Thoughts and feelings can reveal medical conditions.
- **Privacy of Behaviour and Action:** Thoughts and feelings can reveal a certain attitudes which create a foundation for certain social activities.
- **Privacy of Communication:** none
- **Privacy of Data and Image:** none
- **Privacy of Thoughts and Feelings:** reflexive violation
- **Privacy of Location and Space:** none
- **Privacy of Association:** Thoughts and feelings can reveal individual association, e.g amorous feelings for a certain person.

1.3.2.1.1.6 Privacy of Location and Space The Privacy of Location and Space is concerned with one's right to move freely without being tracked and one's right to private places. If this type is violated, following implicit violations are possible:

- **Privacy of The Person:** Frequently visited doctors can reveal certain medical conditions, if such doctors are known specialsts. In general it could imply ill-being.
- **Privacy of Behaviour and Action:** Frequently visited places in general can reveal association and hence implies social activities.
- **Privacy of Communication:** If one's location is knwon, it is possible to intercept (wiretap) one's communication. This also may violate the right to private spaces.
- **Privacy of Data and Image:** If one's location is known, it is possible shoot pictures. This violates the right to one's image (*"Recht am eigenen Bild"*).
- **Privacy of Thoughts and Feelings:** Frequently visited persons may imply certain thoughts and feelings, e.g. having a mistress.
- **Privacy of Location and Space:** reflexive violation
- **Privacy of Association:** Frequently visted places can reveal associations simply by searching in maps or yellow-pages.

1.3.2.1.1.7 Privacy of Association The Privacy of Association is concerned with one's right to associate with whomever one wants, without that association having recorded.If this type is violated, following implicit violations are possible:

- **Privacy of The Person:** Association with tatoo artists could imply having tatoos or other body modifications
- **Privacy of Behaviour and Action:** Association with churches or poltical organizations could imply certain activities.
- **Privacy of Communication:** none
- **Privacy of Data and Image:** none
- **Privacy of Thoughts and Feelings:** Association with churchs of political organizations could imply a certain intellectual attitude.
- **Privacy of Location and Space:** none
- **Privacy of Association:** reflexive violation

1.3.2.1.2 Sensor Privacy Matrix (Figure 9)

1.3.2.1.2.1 Privacy of The Person The **Privacy of The Person** is generally concerned with one could best understand as *Biometric Privacy*. Friedewald et al. paraphase it as *"[. . .] the right to keep body functions and body characteristics [. . .] private"*. **Accelerometer**, **Rotation Vector** and **Gyroscope** measure the physical movement of the mobile device on all three axes. If the moblile device is carried *"normally"* its safe to say that those sensors also measure the movments of its carrier. So his privacy is infringed regarding biometric behaviour, as it is captured automatically. (**Note:** This is not to be confused with the **Privacy of Behaviour and Action** which is used for the social aspects of behaviour action, e.g. praying, sexual habits or poltical activities.)

1.3.2.1.2.2 Privacy of Data and Image The **Privacy of Data and Image** demands, that *"individual's data is not automatically available to other individuals"*. This type of privacy is

trivially threatened because here sensor data is indvidual data, a priori. So every sensor violates the privacy of data and image, as data is transported into a foreign system where operators have access to it.

1.3.2.1.2.3 Privacy of Location and Space According to Friedewald et al., the **Privacy of Location and Space** is concerned with one's *"right to move about in public or semi-public space withoug being identified, tracked or monitored."*. This is the location aspect of this type. The space aspect is concerned with one's *"right to solitude"*, which generally includes one's right to an inviolate home and other private spaces. Obviously the **GPS** and **GSM** sensors violate such right about not being tracked, because they reveal the position of the mobile device and its carrier. The **GSM** sensor gives the exact cell, the mobile device has registered with at the current moment. 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.

The **WLAN** and **Bluetooth** sensors record lists of the currently available local wireless networks and bluetooth clients. If such are known stationary entities, those sensors are considered as dangerous as the **GSM** sensor for the carriers locational privacy.

The **Magnetic field** sensor is not regarded very dangerous to the carrier's privacy, because it does not allow very precise localization. But it can limit the possibilities for the global position of the mobile device. Here, it is just named for completeness sake.

1.3.2.1.2.4 Privacy of Associtation The **Privacy of Association** sates that everyone has the *"right to associate with whomever they whish, without being monitored [automatically without reasonable suspicion]"*. This includes indivisuals and organizations. **WLAN** and **Bluetooth** sensors provide the ability to monitor such associations if their lists contain kown entities. If one frequently connects with an organizational wirelss network, e.g. an university network, an association can be deduced (student or staff). The same goes for the **Bluetooth** sensor, if it is stationary. Additionally, if the recorded bluetooth clients are mobile, it is more or less possible to deduce association with the technical identity of (yet) anonymous individuals.

Additionally, the **GSM** sensor could provide the association with the GSM operator (**NEEDS TO BE VERIFIED!**).

1.3.2.1.3 Excessive Data Mining

1.3.2.1.3.1 1 Privacy of the Person "This type is the right keep body functions and body characteristics private" [1]. Here we are concerned with the personal physiological and psychological privacy of citizens. This privacy is threatened if **body characteristics** (weight, height, body measures, fingerprints, dna,...) or **medical conditions** (limping, having a cold, suffering from depression, visiting therapists or other doctors) **can be revealed or inferred**.

The Live+Gov project collects sensor data from mobile devices (accellerometer, location, ...). By applying *human activity recognition (HAR)* techniques we could detect certain movement patterns and in fact detect limping or other pathologic movements. Additionally we know the position with of indiviual citizens with a sufficient accurancy, and the Live+Gov project is only applied to a destinct urban area. So we could link HAR data with the yellow pages, filter for

medical specialists and limit the possibilities of pathologic conditions. But even without HAR data we could determine a likelihood for certain conplaints. Frequent vistits to the dentist does not imply healthy teeth.

1.3.2.1.3.2 2 Privacy of Behaviour and Action *"This type is also concerned with the 'protection against disclosure of personal matters' through behaviour"* [1]. This type is concered with religous practices, sexaul habits, political activities, etc. revealed through observation. For the Live+Gove project this type is related to the Privacy of Location and Space and the Privacy of Association. Locational data is recorded by default and by linking it to maps and public registers like phone books those *"personal matters*" could be disclosed by frequently visited places (churches, brothels, party head quaters, . . .).

1.3.2.1.3.3 3 Privacy of Communication *"It 'aims to avoid the interception of communications' either electronic or face-to-face"* [1]. This type is threatened if we intrude the secrecy of correspondence, posts and telecommunications, personal direct communication or right to free discussion without third parties listening. The Live+Giv project only could threaten this type by using mobile client application as trojan horse, collecting any communication data (chat, sms, microphone).

1.3.2.1.3.4 4 Privacy of Data and Image *"This type is concerned with 'making sure that individuals's data is not automatically available to other individuals and organisations' "* [1]. This is Informational Privacy in an intuitive sense regarding data like:

- Phone Number
- IP Address
- Public-administrative Data (Date of Birth, population register)
- Data held by organizations, like Banks or Insurance Companies
- All data that is stored in online services (Facebook)

The Live+Gov project does not threaten this type directly if we assume a secure and closed system. However, this privacy type could be threatened by carelessly ignoring known (technical) vulnerabilities regarding data security. Additionally we could threaten the Privacy of Data and Image indirectly by linking collected data with other sources.

1.3.2.1.3.5 5 Privacy of Thoughts and Feelings *"This type is the right 'not to share their thoughts or feelings or to have those thoughts or feeling (sic!) revealed' "* [1]. This means thoughts and emotions must not be detected autmatically. Intuitively the Live+Gov system seems unable to threaten this type. However, considering the issue component of the Urban Maintaince use case, this might reveal information about one's thoughts regarding the community in a positive manner. Solely by taking part we could assume a caring personality. This threatens one's privacy in a rather technical sense, but does not necessarily impose any harm.

An other way of threatening this type could be constructed by recording one's voice with the phone's microphone and run emotion detecting algorithms against this data.

1.3.2.1.3.6 6 Privacy of Location and Space *"This type is the right 'to move about in public or semi-public space withoug being identified, tracked or monitored'. Additionally this type is concerned with the protection of one's home and private places ('right to solitude')"* [1]. The location dimension of this type is relatively easy to understand: The geo-position of citizens cannot be monitored by default. However, by actively taking part in the Live+Gov project, the locational privacy of citizens is threatened by default because data of the location sensor will be collected. The space dimension is more complex, but can be simplified with a *"right to solitude"*. This dimension could be threatend with the invasion of personal space in any means, i.e. by disrespecting one's right to an inviolate home or by undercutting one's comfort zone in an conversation. Live+Gov only utilizes the mobile devices of citizens, so we could disrespect the former by activating the phone's micropone and start recording.

1.3.2.1.3.7 7 Privacy of Association *"This type is the right 'to associate with whomever [one] wish, withoug being monitored'."* [1]. This means that one's associations must not be recoreded by default independet from any suspicion. Anyhow, it does not mean that this right cannot be forfeit given a reasonable suspicion. The Live+Gov project could easily threaten this type just by linking locational data with yellow pages and a map. Even if an association graph cannot be deducted for individual citizens, we could aggregate the data to create an association graph for a whole population as the Live+Gov system is applied to a restricted urban area.

1.3.3 Step 3. Plan Development

1.3.3.1 Security Measures

1.3.3.1.1 The 7 C's of user privacy control (Figure 10)

This is a note on an excerpt from the article *Sociotechnical Architecture for Online Privacy* [1] called **The 7 C's of user privacy control**. Those 7 C's are aspects which should be covered by measures for implementing user privacy. They derive from an interpretation of privacy which could be summarized as *"One's ability to control/seclude information about themselve"*.

1.3.3.1.1.1 Comprehension *"Users should **understand** ho personal identifiable information (PII) is handled, who's collecting it and for what purpose, and who will proccess the PII and for what purpose. Users are entitled to know all parties that can access their PII, the limits to processing transparency, why the PII data is being requested, when the data will expire (Either from a collection or database), and what happens to it after that. This category also include legal rights around PII, and the imolications of a contract when one is formed."*

This C implements transparancy regarding user data and user privacy. Comprehension should answer the following questions:

- *WHO* collects data?
- *WHAT* data will be collected?
- *WHY* will data be collected and processed?
- *HOW* will data be collected and processed?
- *WHEN* will data expire?
- What is allowed?

- What choices ar possible?

All in all information of what's happening and why has to made accessible for users.

1.3.3.1.1.2 Consciousness ***(critical!)*** *"Users should* ***be aware*** *of when data collections occurs, when a contract is being formed between a user and data collector when their PII is set to expire, who's collecting the data, with whom the data will be shared, how to subsequently access the PII, and the purposes for which the data is being collected."*

This C seems to be critical for privacy protection. Consciousness complements Comprehension in respect that the latter just states that hard facts need to be delivered. However, those facts might get hidden in a terms and conditions section which nobody reads but still accepts anyway. In order to prevent that Consciousness states that a certain level of **Awareness** of those facts needs to be established.

1.3.3.1.1.3 Choice "Users should ***have choices*** *regarding data collection activies in terms of opting in or out, whether or not to provide data, andhow to cerrect their data."*

Self explaining. This is the actual control enabled by the 7 C's.

1.3.3.1.1.4 Consnet "Users must first ***consent*** *(meaning informed, explicit, unambiguous agreement) to data collection, use, and storage proposals for any PII. Privacy consent mechanisms should explicitly incorporate the mechanisms of comprehension, consciousness, limitations, and choice."*

This C might be special case of Choice. Before taking part a user should have the choice wether to join or not (Opt-In).

1.3.3.1.1.5 Context "Users should ***be able to change pricacy preferences*** *according to context. Situational or physical context - such as crowded situations (for example, when at a service desk where severeal people can listen in on your exchange when you provide a phone number, or when you're in an online community chat room) - is different from wehn you perform a buy transaction with Amazon.com or in rooms with cameras (where digitization makes the information permanent and unmistakably you) and data context (such as the sensitivity of data, for example health data could dictate different actions on the same PII in different contexts."*

Self explaining. Refines Choice in context sensitive manner.

1.3.3.1.1.6 Confinement "Users should ***be able to set limits*** *on who may access their PII, for what purposes, and where and possibly when it may be stored. Setting limits could provide some good opportiunities for future negotiation between vendors and users."*

Self explaining. Refines Choice regarding data collection and processing.

1.3.3.1.1.7 Consistency "Users should ***anticipate*** *with reasonable certainty what will occur if any action involving their PII is taken. That is, certain actions should be predictable on user acces of PII or giving out of PII."*

Information given by Comprehension needs to be reliable to found choices.

1.3.3.1.2 The 2 Steps of the 7 C's (Figure 11)

If we look closer at the 7 C's and how they try to enable control, we see that a 2 step approach is taken:

1.3.3.1.2.1 (I) Enable Adequate *Control*

The 7 C's try to enable control through choices. A user should be able to choose which data can be collected and processed depending on context and who will have access to the data. But cannot be random. In order to make substantiated choices and enable *adequate* control a user needs have

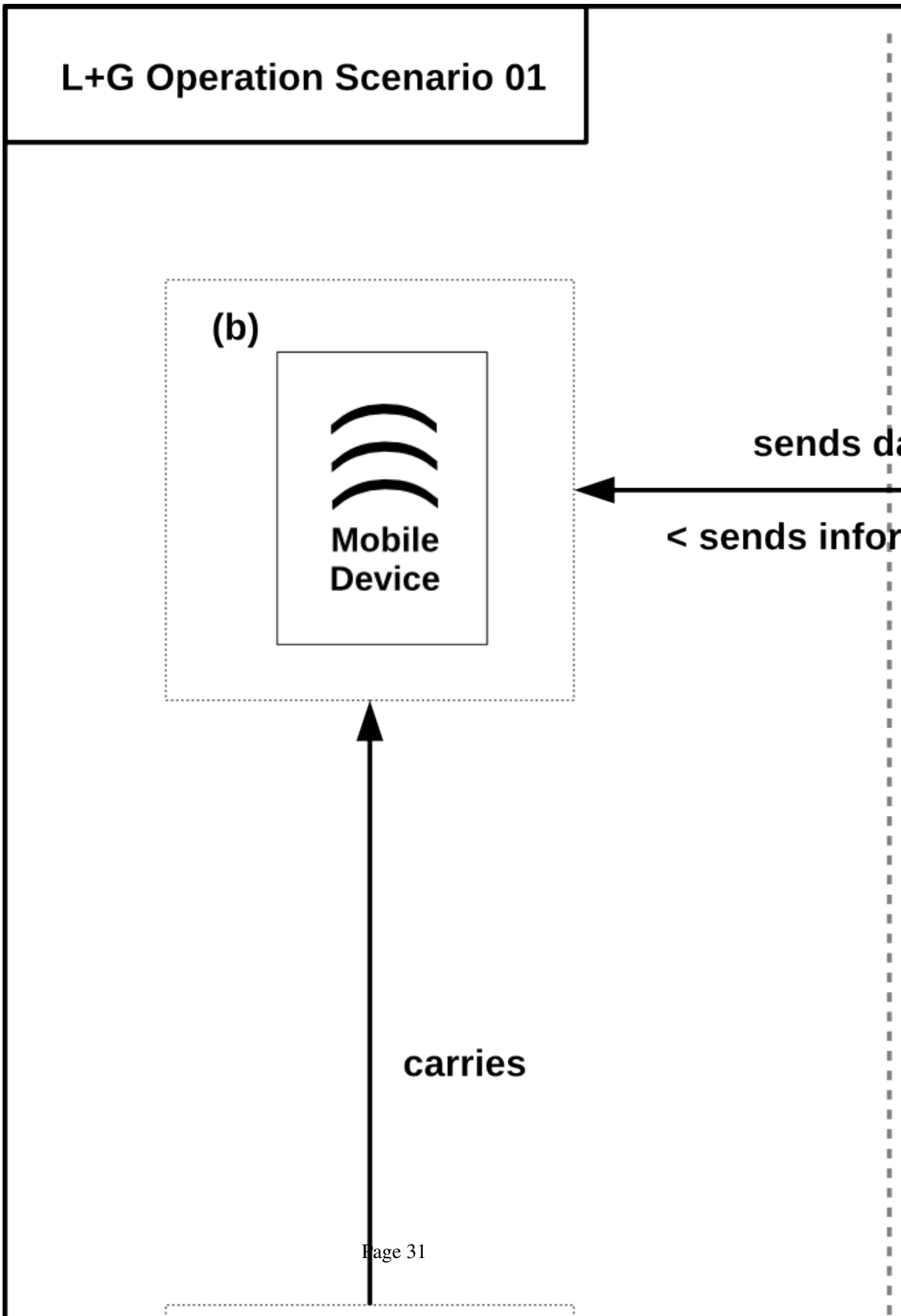
- **Comprehension:** access to hard facts
- **Consistency:** trust that those facts are reliable
- **Conciousness:** awareness of those facts to found choices

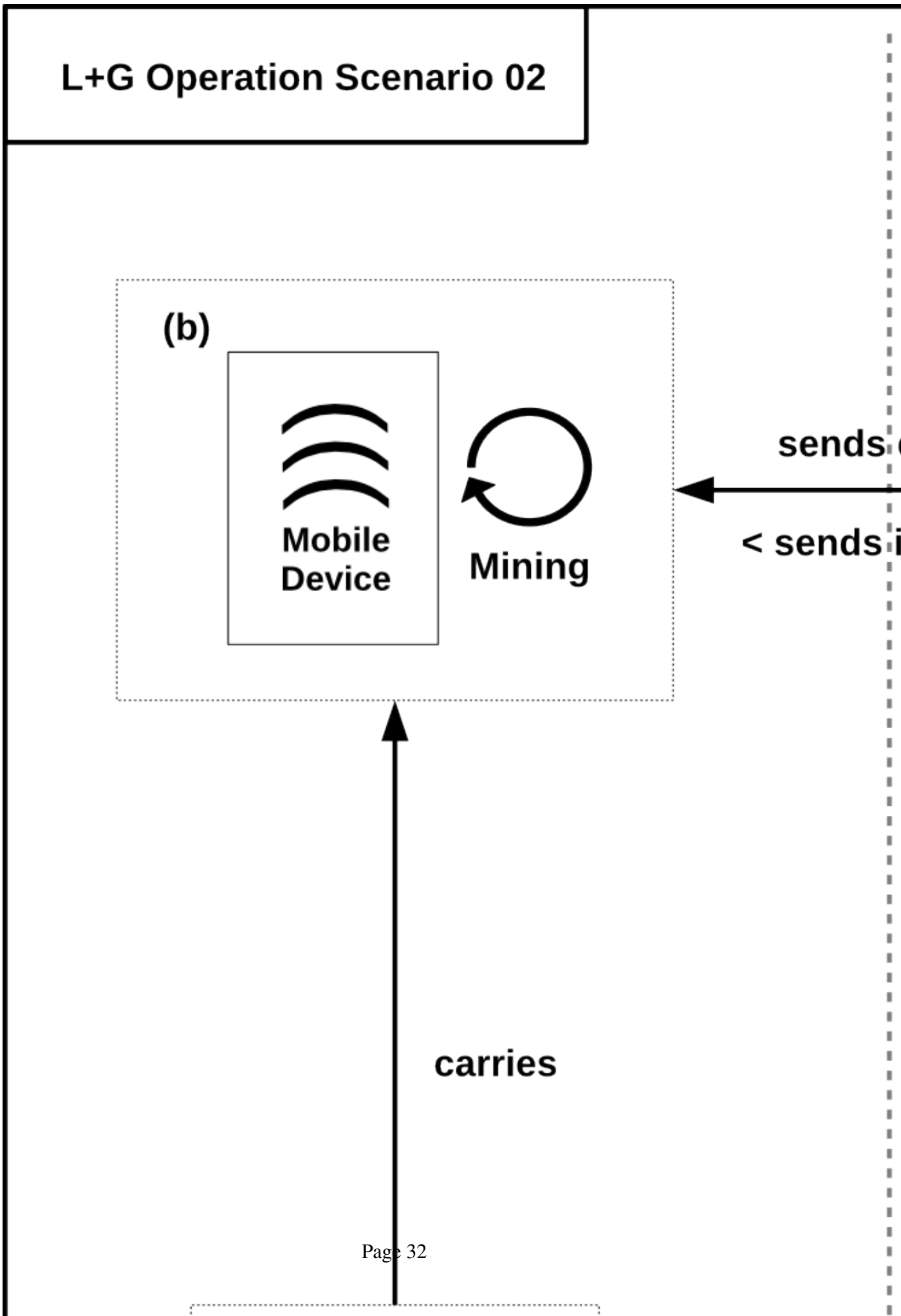
1.3.3.1.2.2 (II) Enable Actual *Control*

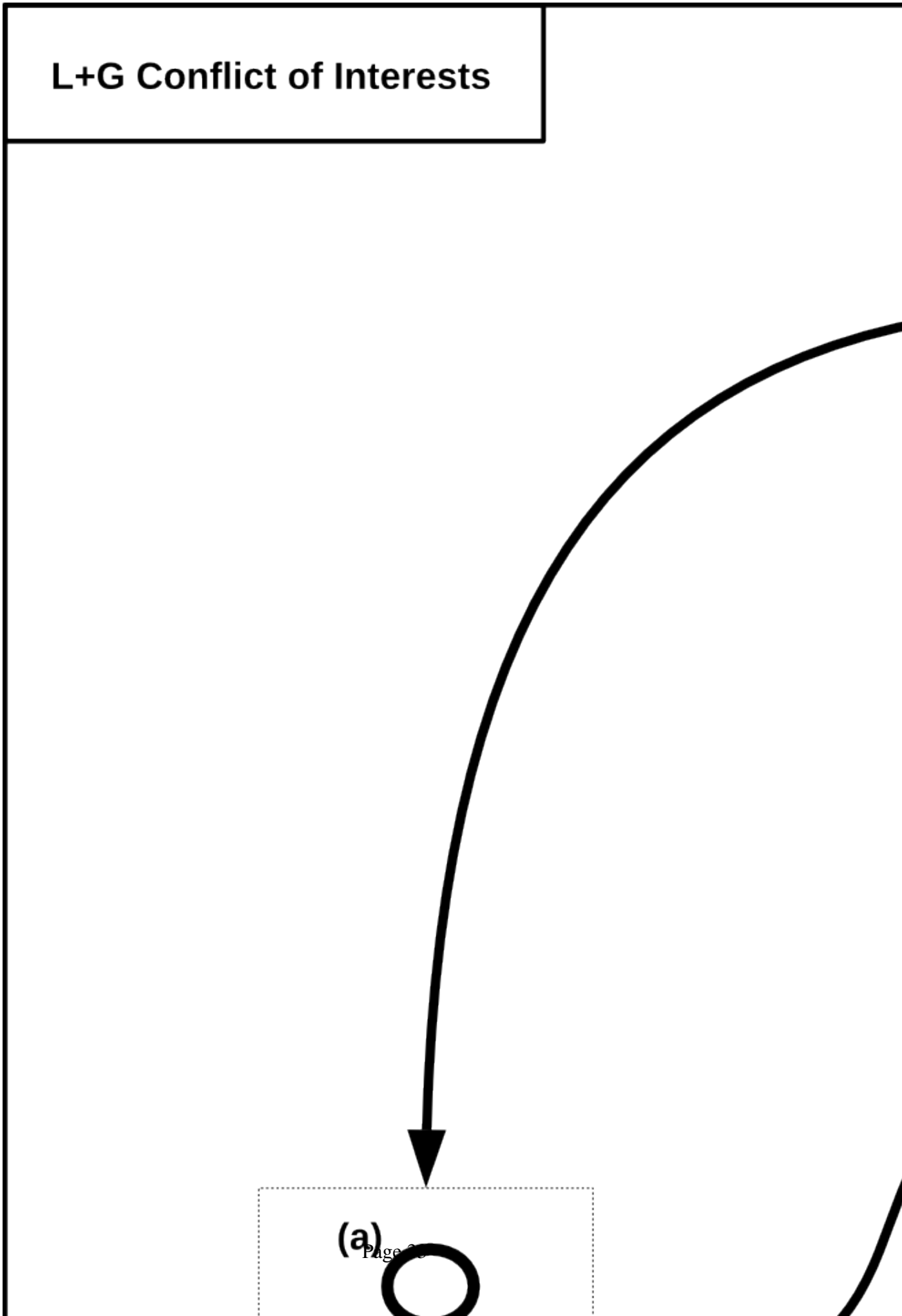
Naturally after creating a certain amount of knowledge, a user needs also access to opportunities to make use of it. Therefore a user needs possibilites to actually make choices. Additionally to having choices at all, the 7 C's have 3 special choice categories:

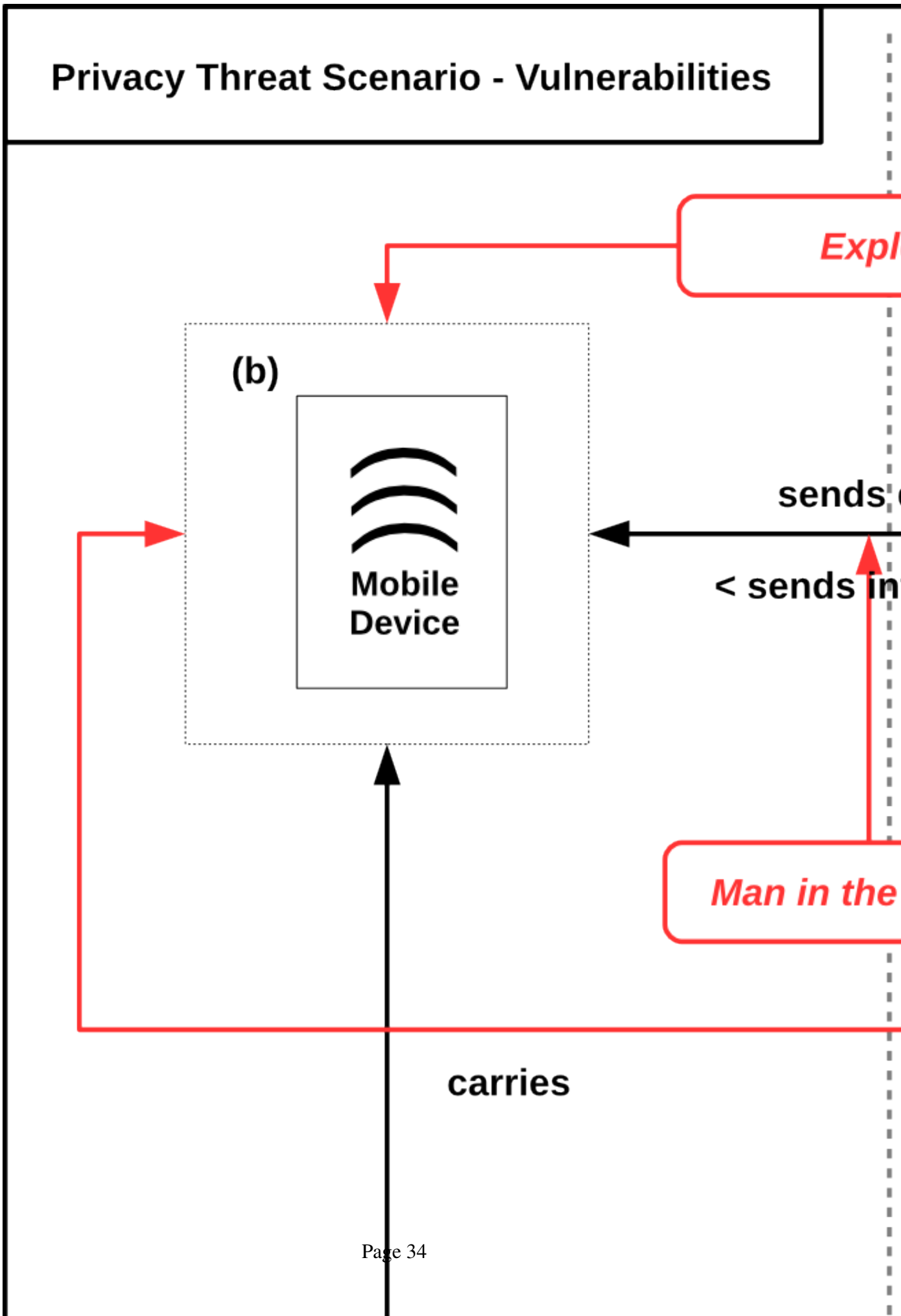
- **Choice**
  - **Consent:** the choice to opt-in
  - **Confinement:** the choice to set limits regarding user data
  - **Context:** the choice to set limits regarding user data depending on certain context

1.4 References


L+G Operation Scenario 01
(b)
Mobile Device
sends da
< sends infor
carries


L+G Operation Scenario 02
(b)
Mobile Device
Mining
sends
< sends
carries


L+G Conflict of Interests
(a)

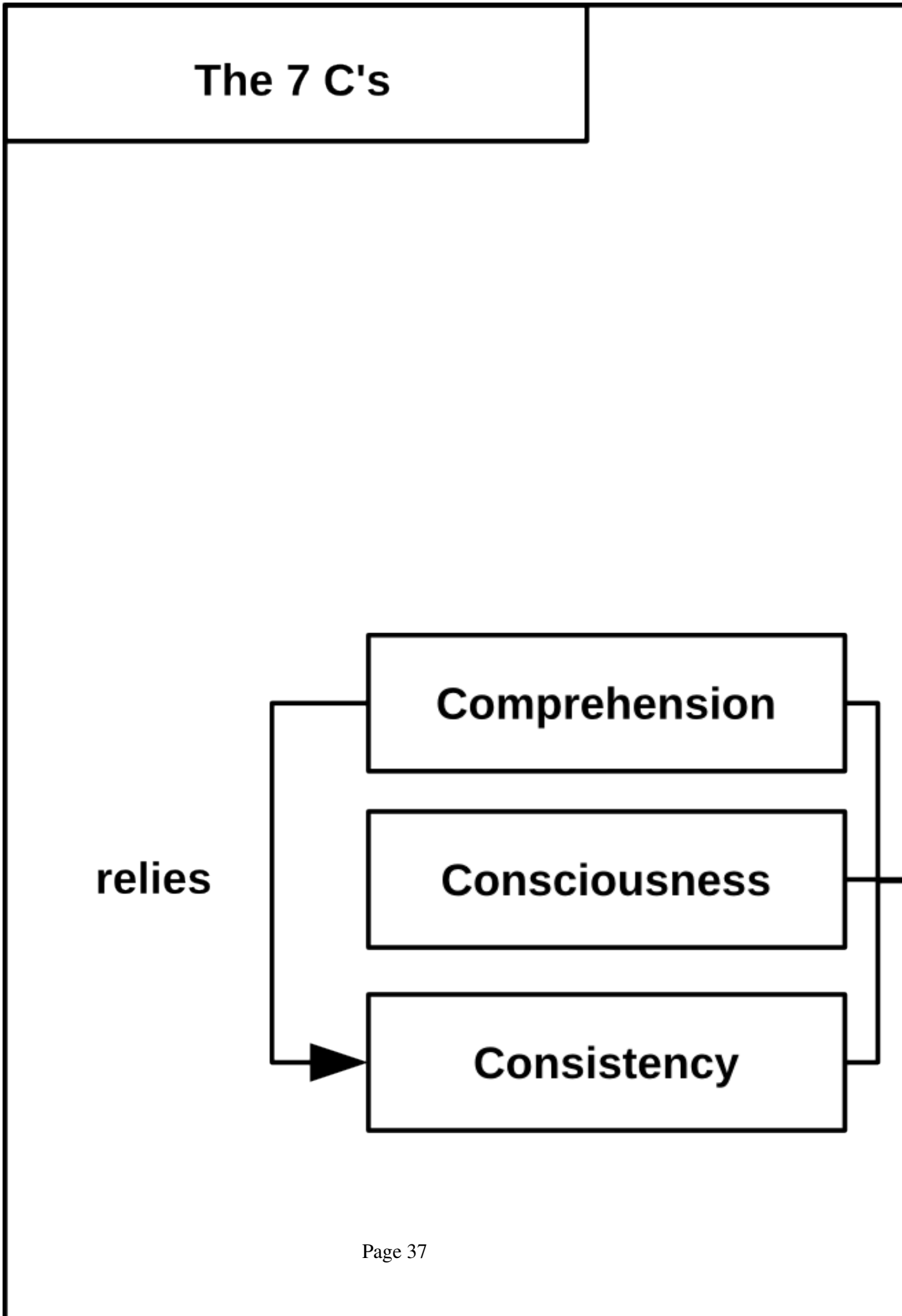

Privacy Threat Scenario - Vulnerabilities
Expl
(b)
Mobile Device
sends
< sends in
Man in the
carries

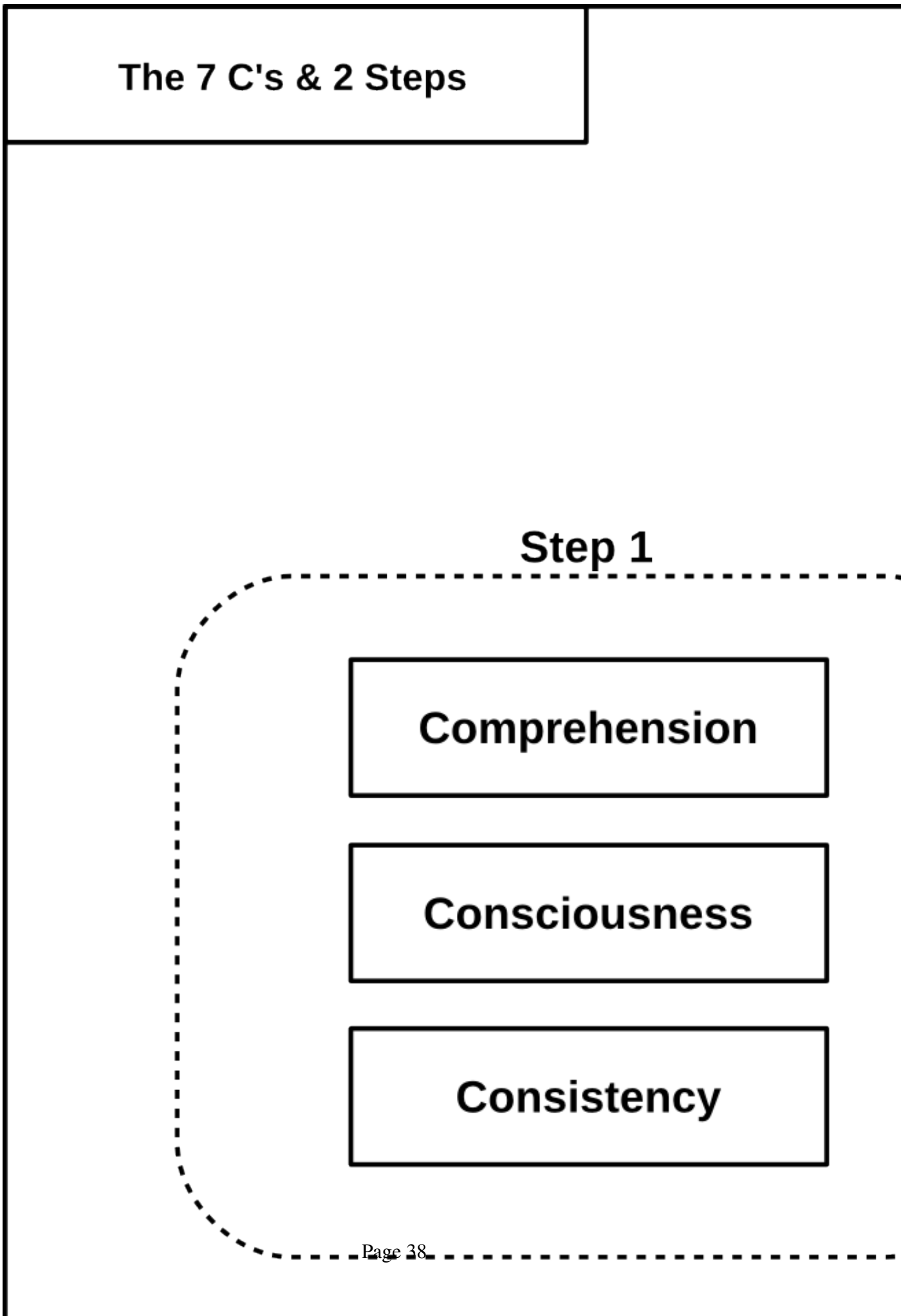
X / Y	Privacy of the Person	Privacy of Behaviour and Action	Privacy of Communic
Privacy of the Person	●		
Privacy of Behaviour and Action	●	●	
Privacy of Communication	●	●	●
Privacy of Data and Image	●	●	●
Privacy of Thoughts and Feelings	●	●	
Privacy of Location and Space	●	●	●
Privacy of Association	●	●	

Implicit Vi

Read: Data gained by violation

	Privacy of the Person	Privacy of Behaviour and Action	Privacy Communic
GPS			
Accelerometer Linear Acceleration, Gravity	●		
Rotation Vector	●		
Gyroscope	●		
Magnetic field			
WLAN			
Bluetooth			
GSM			


The 7 C's
Comprehension
relies
Consciousness
Consistency


The 7 C's & 2 Steps
Step 1
Comprehension
Consciousness
Consistency

1.5 Implementation

1.5.1 Authentification and Access Control

1.5.2 Anonymization and Pseudonoymization

1.5.3 Randomization

2 Improved Sensor Data Mining Methods

2.1 Battery Awareness of Sensor Collector

Results from Projektprakikum (Wifi/Zip)

2.2 Service Line Detection (new method)

MA thesis results from Sven Milker

2.3 Issue Analysis

Christoph Schfer with Niko Beck