iKoPA Integrierte Kommunikationsplattform für automatisierte Elektrofahrzeuge

Deliverable 3.3

Data Protection Impact Assessment of Mobility-Related Communications

Version number	1.0
Dissemination level	Public
Project Coordination	htw saar
Due date	2018-11-30
Date of preparation	2018-12-17

Funded by the

Federal Ministry
of Education
and Research

Project Coordination

Prof. Dr. Horst Wieker

Head of ITS Research Group (FGVT) at the
htw saar – Hochschule für Technik und Wirtschaft des Saarlandes,
University of Applied Sciences
Department of Telecommunications
Campus Alt-Saarbrücken
Goebenstr. 40
D-66117 Saarbrücken
Germany

Phone +49 681 5867 195
Fax +49 681 5867 122
E-mail wieker@htwsaar.de

Author:

Bud P. Bruegger – ULD

Revision and History chart

Version	Date	Description
0.9	06/12/2018	**Final draft for review**
1.0	17/12/2018	**Version edited based on review comments**

Table of Contents

Figures

Tables

Executive Summary

iKoPA technology has been developed by following a process of data protection by design. This has resulted in an architecture that mitigates data protection risks as much as possible with the current state of the art. The residual risk of iKoPA technology has therefore been found to be small.

The present document describes the methodology that was developed for the assessment of iKoPA technology. This includes the identification of key risks and a systematic presentation of the concepts behind the current state of the art of mitigation measures. A study of the effectiveness of these measures and the identification of areas where they may be further improved contributes to future research aimed at pushing the state of the art of data protection.

The experience in iKoPA has shown the importance of taking into account all communications happening in the wider context. The impact assessment methodology of iKoPA is therefore applied beyond iKoPA technology to the wider context of mobility-related communications. An initial assessment has already identified significant areas with yet unmitigated and at times considerable risks. This confirms the reusability of the developed concepts and methodology. The initial success also invites future research that applies the iKoPA methodology more systematically and thoroughly.

The impact assessment focuses on the risk of collecting location data of identified persons and the collection of such data into movement profiles. It has analyzed the mitigation measures that are common in mobility-related communications, namely pseudonymization, frequent and coordinated change of pseudonyms, and separation of domains. The analysis includes an assessment of the effectiveness and limitations of these measures.

On this basis, the severity of the residual risk of mobility-related communications after mitigation is being discussed. It found that a multitude of highly motivated and capable attackers exist, that attacks are relatively simple to deploy and hard to detect, and that even at this point of time, a systematic attack that collects movement data of modern vehicles is under way. It is important to note that this attack does not involve iKoPA technology.

The report therefore concludes that currently, the residual risk in mobility-related communications is very considerable and recommends that urgent actions be taken to address the current shortcomings. The following is an incomplete list of findings:

- The biggest danger stems from the current systematic large-scale collection of vehicle location data most likely in a third country. This is made possible by WiFi access points that are built into vehicles and lack any mitigation measure such as pseudonymization, together with the existing smartphone location services by Google and Apple.

- The current generation of privacy-enhanced RFIDs still lacks the possibility of assigning changing pseudonymous TAG identifiers. An improved generation of RFIDs needs to be developed before wide-spread application in a mobility setting can be considered.
- The combined use of a multitude of communication technologies and channels even within the same transactions raises new risks. The mitigation measure of pseudonym change is potentially less effective since a coordinated change of pseudonyms is often not possible across multiple channels. How severely this reduces the effectiveness of this mitigation measure is not yet fully understood. Therefore, additional research is necessary to address this issue (see below). It is therefore also recommended that data protection working groups of C-ITS also address the wider scope of mobility-related communications.

1 INTRODUCTION

This section starts with describing the application of data protection by design in the iKoPA project that resulted in a state of the art solution and as part of which the methodology described in this report was developed. The section then describes the target of evaluation to which this methodology is applied to, namely the entirety of mobility-related communications—an artifact that goes far beyond iKoPA technology and use cases. Previous work on impact assessments in the context of mobility-related communications is reported in subsection 1.3. Subsection 1.4 lists the risk factors that were extracted from a literature search in the area of linkability and tracking. A limitation of scope then sets the expectations for the present assessment. The section concludes with an outline of the remainder.

1.1 Data Protection in iKoPA

iKoPA has followed the approach of **data protection by design** by making data protection an integral part of all project phases. This was facilitated by having a data protection authority, namely the research section of the Unabhängiges Landeszentrum für Datenschutz (ULD) of Schleswig-Holstein, as project partner.

This approach included the training of technical partners in the concepts of data protection in data protection workshops, awareness building of privacy-critical issues, joint identification of possible risks, consulting on arising privacy questions, and the study of available privacy enhancement technology and state-of-the-art mitigation measures. The development of technology was guided by a total of 56 formal privacy requirements (see section 5.2.1 in [iKoPA D1v2 2018]). They were structured in six categories (namely, availability, confidentiality, integrity, intervenability, transparency, and unlinkability) and already at their conception considered the new General Data Protection Regulation (GDPR). A later evaluation (see [iKoPA D4v2 2018]) showed that all the requirements were taken into account and that no unjustified deviations were found.

In only one case, an only partly mitigated risk was identified in iKoPA. Namely, even the current state of the art in radio frequency ID (RFID) technology fails to support the change of identifiers. This shortcoming could only be partly mitigated through the use of trusted computing technology (see [iKoPA D5v2 2018], section 3.2.6). The need for an improvement of privacy-protection in RFIDs was therefore documented (see [iKoPA D1v2 2018] section 4.3.3.3.2.). iKoPA also supports an alternative that uses vehicle2x instead of RFIDs technology and thus avoids the described shortcoming.

Apart from this single unavoidable issue, all the technology developed in iKoPA can be considered to represent the **current state of the art of data protection**[1].

The present document describes among others the **methodology to assess data protection impact** that was developed in iKoPA. The methodology was used in iKoPA to identify and successfully mitigate data protection risks. It is now also **applied to a wider context** around iKoPA where it has already successfully identified significant, yet unmitigated risks (see section 6 below).

It is important to note that the risks identified in a wider context are **not** shortcomings of iKoPA. For example, one of the most severe risks stems from built-in WiFi access points of vehicles that are not part of iKoPA technology or even used in iKoPA. Other risks originate in the fingerprinting of web browsers. Also web browsers are not part of iKoPA technology or incorporated in any iKoPA use case.

In addition to applying the iKoPA methodology to a wider context, this document also provides a conceptual description of the major state of the art mitigation measures that were used in iKoPA (see section 5 below). Part of this description is an assessment of effectiveness. Understanding the limitation in the effectiveness of measures is an important first step for advancing the state of the art of data protection technology. As a contribution, this document has revealed possible difficulties in the coordination of pseudonym change across multiple communication channels (see section 5.2.3.3. below). This may trigger future work that studies this potential issue in further detail and identifies strategies to improve the currently known mitigation measures and thus push the state of the art.

1.2 Target of Evaluation: Mobility-Related Communications

This section describes the target of evaluation, i.e., the technical artifact whose impact on society is being studied in this assessment.

The iKoPA project is situated in the context of Cooperative Intelligent Transport Systems (C-ITS) and Cooperative, connected and automated mobility (CCAM) [EC C-ITS Platform]. Privacy and data protection considerations of C-ITS have already received ample attention (see section 1.3 below for detail). The focus of these studies lies on the communications of a vehicle with other vehicles and with road-side stations. This is typically called vehicle2x communication and includes Cooperative Awareness Messages (CAM) and the Decentralized Environmental Notification Messages (DENM).

[1] Note that even though iKoPA technology represents the state of the art, every application of the technology must again take data protection into account. The sole use of the technology fails to guarantee compliance with the GDPR.

Work on iKoPA has shown that a study of the privacy impact needs to address a wider context. This is for example illustrated by the requirement "REQ-PUL-009 Unlinkability of multiple reservations" (see [iKoPA D1v2 2018], section 5.2.1, page 198). Here, the reservation is first booked from a smartphone over a mobile IP network (see Use Case UC-01 in [iKoPA D1v2 2018], section 3.3.3.1). The resulting ticket is then presented by the vehicle to the entry barrier of a parking lot over vehicle2x communications (see Use Case UC-04 in [iKoPA D1v2 2018], section 3.3.3.4). Evidently, looking only at the unlinkability of the involved vehicle2x communications is insufficient, since linking is also possible across consecutive reservations from a smartphone over the mobile IP network.

This extension of the focus to all involved communication channels in iKoPA has been further expanded for the target of evaluation of this impact assessment to include all communications that happen in the context of mobility. In addition to communication channels actually used in iKoPA, this includes also metadata from WiFi access points built into vehicles and potential communications from a web browser with some reservation service.

This can be summarized by stating that the **target of evaluation** of this impact assessment is the **entirety of communications that happen in the context of mobility**.

1.3 What Risks have already been studied?

This section briefly reviews the data protection issues that have been studied so far in C-ITS and CCAM. This prepares for the following section that describes the additional risk factors to be considered for assessing the entirety of mobility-related communications. What is of particular interest here is the definition of the respective targets of evaluation of the existing studies. This section illustrates in particular the predominant focus on vehicle2x communications.

(i) The EC has organized its activities in the area of "Cooperative, connected and automated mobility (CCAM)" in the C-ITS Platform that is divided in two phases covering 2014-2016 and 2016-2017, respectively (see [EC C-ITS Platform]). The work has been conducted by 10 active working groups, all chaired by DG MOVE representatives in cooperation and with active participation of other Commission services (see [EC 2016a] at the bottom of page 17. Since data protection and privacy are important design considerations, the working group 4 is "Data protection and Privacy" (see [EC 2016a], section 6, pages 48-60). The activities of this working group are reported in [EC 2016a] section 6 and [EC 2017] section 4, pages 27-32.

For phase I, an excerpt of the executive summary of working group 4 in section 6.1 of [EC 2016a] illustrates the focus of work (see page 48):

"To test and develop solutions, WG4 identified a family of standardized messages relevant for the purpose of analyzing the privacy and data protection challenges faced by C-ITS: the Cooperative Awareness Messages (CAM) and the Decentralized Environmental Notification Messages (DENM)."

In other words, the study has focused on a **single communications channel**, namely vehicle2x (i.e., a variation of WiFi for wireless access in vehicular environments**,** see for example [Wikipedia IEEE_802.11p]) and its CAM and DENM messages.

For phase II, section 4.1 on page 27 of [EC 2017] states that "The objective of the C-ITS Platform Working Group on Data Protection & Privacy in phase II was to conduct an analysis of the suitable legal bases for lawfully processing personal data and to analyze the deliverables of phase I against the background of the General Data Protection Regulation (hereinafter GDPR)." It further states on the same page: "The aim of the working group was to obtain guidance and points to be taken into account from Article 29 Working Party in order to reach a sound level of protection of personal data also in relation to the GDPR. The document 'Processing Personal Data in the Context of C-ITS' was submitted to the representative of the technology subgroup of Article 29 Working Party on the 10th of July 2017 with a view to receive an opinion in the beginning of October 2017."

In this work, *tracking* has been identified as one of the risks and *certificate change strategy* identified as a major control for its mitigation (see section 4.3.2 on page 30). On page 31 it states that "Furthermore, the working group sees that from a technical perspective key elements of data processing using privacy by design have been incorporated from the initial design of the C-ITS system: limited range and radius (300 meters), alternating pseudonym certificates to avoid tracking, local 'on the spot' exchange of data between neighboring vehicles with little data retention and no further processing."[2] In the technical recommendations in section 4.4.3 on page 32 it further states: "The working group is of the opinion that further mitigation measures concerning the possibility of tracking should be taken, such as analyzing how static data in CAM can be used on their own or in combination with other information to identify a single vehicle as well as analyzing any appropriate type of 'do not track' functions, as well as encryption."

This means, that also in phase II, the focus of work was on the same **single communications channel** of vehicle2x.

(ii) In response to the request by the data protection working group of the C-ITS Platform in phase II (see above), the Art. 29 Working Party issued an Opinion on Processing personal data in the context of Cooperative Intelligent Transport Systems (C-ITS) [Art29WP 2017]. The description of the processing activities can be found in [C-ITS Platform DPWG 2017]. On this basis, the Art. 29 WP summarizes the personal data that

[2] Note that other authors use different values for the range. For example, [Ullmann 2017b] (" in section III on page 36) estimates a "range of up to 800 m in open space".

are processed (see section 3.1 of [Art29WP 2017] on page 4). They consist of the CAM and DENM messages transmitted via vehicle2x and the message authorization certificates used for these transmissions.

This summary illustrates that also the Art. 29 WP has only focused on the **single communication channel of vehicle2x**.

(iii) To guarantee the legitimacy of CAM and DENM messages sent over the vehicle2x communication channel, pseudonymous authorization certificates are used in an open readable form (namely in an unencrypted broadcast or geocast message). A major data protection concern is the tracking of persons based on such pseudonym certificates. A study by ETSI [ETSI 2018b] therefore studies various options of pseudonym change management and how they can mitigate the tracking risks.

Again, also the ETSI study focuses on the **single communications channel of vehicle2x**.

(iv) [Eisses et al 2012] discuss personal data protection in the context of ITS, taking into account a wider range of applications than those foreseen for Day1 and Day1'5 (see page 4). In section 3 (pages 50 through 110) they analyze the privacy aspects of each application, including the applicability of various protection measures. In particular, the following measures were considered (see page 54 for a list and pages 6-7 for a description of some of these measures). The order of the list was changed by the author to illustrate a point.

1. anonymization
2. pseudonymization
3. domain separation
4. distributed processing
5. data minimization
6. deletion immediately after initial processing
7. user consent mechanisms
8. data subject control

In the context of this impact assessment, it is interesting to notice that at least the first four measures of this list, and potentially also measures 5 and 6, control **linkability** of data sets. Also, Table 4 (see page 110) illustrates the importance of measures to control **linkability** for the majority of applications.

While some of the analyzed applications (such as road user charging) utilize other communication channels than vehicle2x (for example, SMS), the risks inherent in the use of multiple channels in a single transaction fall outside the scope of the study.

(v) ETSI's technical specification on "Trust and Privacy Management" of the ITS [ETSI 2018a] describes the main approach to privacy in section 5 (bottom of page 10). In particular, it states that anonymity is unsuitable for the ITS and that observability is required for safety. It further states "Consequently, pseudonymity and unlinkability offer

the appropriate protection of the privacy of a sender of basic ITS safety messages (CAM and DENM)."

It is evident that the specification regards solely the single communications channel of vehicle2x (with its CAM and DENM messages). It also underlines the importance of linkability control (for example through the use of pseudonyms) for the privacy in the ITS.

In summary, the previous work usually looked into the risks in situations where predominantly the single communication channel of vehicle2x was used. The importance of unlinkability has also been underlined.

1.4 Risk factors related to linkability and tracking

To better understand what risk factors are relevant, a literature search was conducted that also looked beyond mobility. This provided the input to identify five risk factors to consider. The following first describes the situation in which risk is studied and then summarizes the identified factors.

Section 1.2 above has described how in the target of evaluation, even a single transaction with a complex service (such as reservation of a parking spot or charging station) can be composed of multiple communication steps, based on different technologies, and using different channels.

A communications channel typically consists of a stack of layers (see for example [Wikipedia OSI_model]). Each layer may use some identifier. For example, while the application itself may use some pseudonym, lower levels of the stack may employ their own identifiers such as a MAC address at the link access layer or an IP address at the network layer. Typically, different stakeholders in the communication (recipient, network operator, and eavesdropper) have visibility of a different subset of these identifiers.

Further, when a transaction consists of multiple communication steps, possibly over different communication stacks, a transaction identifier (or pseudonym) is typically necessary to be able to link the steps into a single transaction.

As has been discussed above, linkability is a major concern in the context of ITS. The basic idea in the above mentioned literature is therefore to achieve unlinkability by avoiding the use of the same identifier in distinct transactions. Typically this is achieved by using a pseudonym whose use is restricted to a single transaction.

While this approach is evidently effective in simple settings, the situation is rendered far more complex by the use of multiple communications stacks (each with multiple identifiers) for a single transaction. Also, linkability may be further influenced by the context of an application.

The mentioned literature search to help identify the risk factors that apply to this situation yielded examples of risks. The following is an attempt at a systematic compilation of identified risk factors:

(i) **Identifiers must be changeable**: It must be possible to **change identifiers** before starting a new transaction. There are however communication technologies commonly used in the context of C-ITS where this is not currently possible. Namely, even the latest generation of RFID technology (see [iKoPA D1v2 2018] section 4.3.1.4) is unable to change its identifiers (Tag ID and Electronic Product Code) between transactions. Similarly, a vehicles number plate that can be automatically recognized by cameras represents a long-term identifier.

(ii) **All identifiers of a single stack need to be changed in coordination**: Linking can only be prevented, when all identifiers at different layers of a given single stack are changed between unlinkable transactions in a coordinated fashion. While this seems to be current practice in the vehicle2x stack (see for example [Whyte 2016] slide 23), an application-initiated change of IP address in a mobile network stack may be much more problematic[3]. Similarly, when an application can be accessed over a web browser, the avoidance of a reoccurring browser fingerprint may be difficult. Also, if a given stack is used to engage in multiple independent transactions concurrently, the end of one transaction may fall in the middle of another transaction. This renders it more complex to find suitable strategies for changing identifiers across the stack.

(iii) **Multi-stack applications need to coordinate identifier changes across stacks**: When a transaction involves multiple communication steps over different stacks, then unlinkability requires a coordinated change of all identifiers across all stacks. To the knowledge of the author, strategies of pseudonym change have so far been discussed only for a single stack at a time (see for example [ETSI 2018b]). This raises the question with what measures unlinkability across multiple stacks can be achieved and what risks arise from the use of multiple stacks in a single transaction.

(iv) **Quasi-identifiers in the data content may enable linkability**: Even when it is possible to find a strategy to avoid linkability of unique identifiers across distinct transactions, the data sent as part of a transaction may have the character of a pseudo-identifier. Most notably this is the case for

[3] Note that when network address translation is used, the TCP timestamp can still be used as a quasi-identifier (see risk factor (iv) below).

time/location data sent in broadcasts in CAN and DENM messages. This possibility of linking was already pointed out by the Art. 29 Working Party (see [Art29WP 2017], second risk of re-identification, page 7) and was studied in detail in [Ullmann 2017b]. The same risk may also occur with other pseudo-identifiers used in different communication stacks, for example, browser fingerprints or TCP timestamps (see for example [Silbersack 2005] page 6, right column, 2nd paragraph).

(v) **Identity beacons**: Even when an application is designed to achieve perfect unlinkability, the **context** in which this application operates may enable linkability or even direct identification of the data subject. In particular, this is the case, when "identity beacons" advertise the subjects' identifiers in the clear over the same channel or in an otherwise linkable manner (for example by matching of location/time). Englehardt et al describe web servers that act as identity beacons [Englehardt 2015]. Vanrykel et al observe the same beacon behavior by smartphone apps [Vanrykel 2017]. While these beacon identities may be only visible to advanced attackers, modern vehicles come with easier to obtain identity beacons such as the visually observable number plate or the Bluetooth or WiFi MAC Addresses or SSIDs of the infotainment system. The former is used in applications such as road toll (see for example [Nikel 2018]) or parking (see for example [Parkeon 2018]). Ullmann et al studied the kinds of secondary vehicle identifiers introduced through the now common communication technology [Ullmann 2017a]. Tracking of vehicles based on the Bluetooth of passengers' smartphones is described in [Ali 2016]. [TrafficCast International 2018] and [Traffic Solutions Inc. 2018] describe roadside units for the acquisition of Bluetooth and/or WiFi Mac Addresses or SSIDs, respectively.

This list of issues illustrates that the question of reaching unlinkability through the use of pseudonyms is much more difficult to understand in a complex multi-stack environment and taking into account the context in which transactions are performed.

A recent research theme analysis report on cooperative intelligent transport systems by the EC [EC 2016b] states (section 3.2.1 on page 28) that:

> *"Data protection, privacy and security are requirements of most ITS projects. However, so far, most projects have addressed these issues in an ad-hoc/proof-of-concept manner resulting in a limited understanding of the core issues which need to be addressed across Europe."*
>
> *"Further analysis is required to assist in defining common methodologies that address issues of data protection and security among transport projects."*

The present impact assessment therefore attempts to study the question of linkability in complex scenarios systematically and identify both risks and possible mitigation measures.

1.5 Limitations of Scope

This section delineates between what is part of this assessment methodology and what is not. In particular, vehicle data that is not communicated is excluded, this document fails to contain a data protection impact assessment according to paragraph 35 of the GDPR, and of the six common protection goals, only unlinkability is considered.

1.5.1 Exclude data from vehicle to manufacturer

Modern vehicles collect vast amounts of data that are typically collected by vehicle manufacturers and are of significant business value for other parties (see for example, [IWGDPT 2018]). The present study excludes these data from its scope and concentrates on the data that is communicated as part of transactions in the context of mobility.

1.5.2 Exclusion of GDPR Article 35

A data protection impact assessment (DPIA) can have a wide range of forms and scopes. This section attempts to clarify what to expect from this data protection impact assessment. It does so by comparing it to that defined in Article 35 of the EU's General Data Protection Regulation [GDPR 2016]. The comparison is illustrated in Table 1.

Table 1: Comparison of the DPIAs according to Article 35 of the GDPR and that of a new technology, respectively.

	DPIA according to GDPR Art 35	**DPIA of a new technology**
Target of Evaluation	Data processing pertaining to a single purpose as operated by a single controller or exceptionally a small group of joint controllers.	An ecosystem for an open number of data processing activities by an open number of controllers for a wide range of possible purposes.
Known Aspects	• Purpose • Controllers and Processors • processed personal Data • Systems and Processes	• Character of the new technology • Some possible application scenarios

	• Employed protection measures	
Key Questions of the DPIA	• Is the purpose/processing supported by legal basis? • What are the risks for the rights and freedoms of data subjects? • Do the employed protection measures mitigate these risks sufficiently?	• What new risks are introduced by the new technology and its possible applications? • How severe are these risks? • How effectively can they be mitigated with existing or new protection measures?
Systematic Approach	Application of the full set of 6 protection goals	Only protection goals that are significantly affected by new technology and its possibilities and whose analysis is supported by known circumstances.

The table illustrates the vastly different character of the two kinds of DPIAs. In particular, the DPIA according to the GDPR analyses a highly concrete setting of a single processing activity while a DPIA of a new technology addresses a much vaguer defined ecosystem with a wide range of partly difficult to foresee possible processing activities.

First proposals for a systematic methodology to tackle DPIAs according to the GDPR are emerging (see for example [Friedewald 2018] including its review of alternative approaches in section 2.4, pages 10-13). Some of these are based on the Standard Data Protection Model (see [Rost 2002][Rost 2017][SDM-de 2018][SDM-en 2016]). In turn, the systematics of the SDM originates in a complete set of protection goals (see [Rost 2009][Rost 2011][Hansen 2015]). They are:

- Confidentiality
- Integrity
- Availability
- Unlinkability
- Transparency
- Intervenability

The former three protection goals are already known from computer security; the latter three are specific to privacy and data protection. Systematic application of these six protection goals leads to a comprehensive coverage of data protection concerns in DPIAs according to the GDPR.

This poses the question, whether the same systematics can be used in DPIAs of new technologies. Generally, this is not possible, since the evaluation of some protection goals requires concrete settings. These are often simply unknown for a new technology since it depends on concrete decisions of how this technology is used for a particular processing activity and purpose. Since the technology usually leaves a wide range of options open, any analysis in this direction could be conducted on a purely hypothetical basis. The result would thus only apply to an assumed scenario of use and be of little or no value to actually understand the impact of the new technology better.

A DPIA of a new technology should thus only consider those protection goals that can be analyzed without the need of arbitrary hypothesis and assumptions. For this reason, a DPIA according to Article 35 of the GDPR is excluded from the scope.

1.5.3 Restriction of protection goals under consideration

This section looks at the six protection goals and discusses which of them can be analyzed for the new technologies in the context of mobility-based communications.

Unlinkability:

It was already discussed above in sections 1.3 and 1.4 above, how the newly introduced data protection risks in mobility-related communications are closely related to pseudonymity and its use to achieve unlinkability of distinct transactions of a single processing activity and across processing activities by different controllers that serve different purposes. Unlinkability is also the concept that controls the identification of a data subject since identification means to link to a well-established long-term identifier.

As will be reasoned later, linkability can be understood in terms of matching different kinds of identifiers across otherwise disjoint data sets. Since the technology mobility-based communications directly introduces different kinds of identifiers, unlinkability is a property that can indeed be studied.

Confidentiality

Mobility-based communications can use different communication channels and technologies. Different channels and communication styles (such as unicast, broadcast, or geocast) exhibit different confidentiality properties. In particular, depending on the communications technology used, different stakeholders such as intended recipient, network operator, and eavesdropper have access to a different subset of data and

metadata. The protection goal of confidentiality is therefore directly relevant for the DPIA of mobility-based communications.

Confidentiality in the sense of data visibility by different stakeholders is closely related to unlinkability. This is evident when considering that a given stakeholder can only link two distinct data sets if he has the visibility of at least one common identifier in both data sets.

Beyond communications between architectural components, confidentiality is also relevant in the implementation of each single component. For example, a service provider implementation shall restrain access to personal data according to what is needed to fulfill the purposes of the data processing. Without a set of concrete purposes or systems, this aspect of confidentiality could be treated purely on a hypothetical level and is therefore excluded from the scope of this DPIA.

Integrity

In communications, message integrity is an important issue. Each of the considered communications technologies uses different means (such as message authentication codes or electronic signatures) to guarantee integrity. What is new in a mobility context is that multiple channels are combined in a single transaction. The combination of channels does not introduce additional issues of message integrity, however. Since message integrity is thus not related to newly introduced risks, it is excluded from the scope of this DPIA. The same goes for integrity issues in the implementation of architectural components that could solely be discussed on a hypothetical basis.

Availability

By decoupling services from particular communications channels and thus supporting multiple alternative channels for a given message (as is done in iKoPA), the overall availability of services can be improved. It does not however introduce any additional issue to availability. Availability related risks can thus only stem from the implementation of architectural components which is beyond the scope of this DPIA.

Transparency

Transparency mandates that data subjects can understand who processes their personal data for what purposes. Mobility-related communications technology is agnostic of who employs it for which purposes. Risks stemming from lack of transparency could thus only be discussed on a purely hypothetical basis. They are excluded from this DPIA.

This said, this DPIA attempts identify the risks of linkability in use scenarios rendered possible by the new technology. The DPIA in itself can thus be seen as contributing to transparency by increasing the understanding of the data protection issues introduced by the new technology.

Intervenability

The possibility of intervention by data subjects (as parts of their rights extended by the GDPR) is an issue of the implementation of architectural components. It strongly depends

on the actual processing activities. Since the new technology under assessment supports a very wide range of possible processing activities, any evaluation would be purely hypothetical and this protection goal is therefore excluded from the scope of this DPIA.

In summary, this DPIA is solely concerned with unlinkability of different communication steps and the confidentiality of different data and metadata relative to different stakeholders.

1.6 Outline of the Document

The remainder of this document is structured as follows. The next section describes the risks that will be assessed. Then, the basic concepts are defined to describe identification and linking. To better understand the target of evaluation, a model of communications is then presented. On this basis, the concepts of the state of the art mitigation measures for the considered risks are described and their efficiency is assessed. On this basis, the residual risk of mobility-related communications after mitigation is estimated. The conclusions then give a summary of the situation and make recommendations for future actions.

2 RISKS TO BE ASSESSED

While section 1.4 above has identified factors that influence risks, this section identifies the risks that will actually be assessed for the technology of mobility-related communications. For this purpose it looks at how the situation has changed from the times when mobility-related communications were absent to the complete roll out of this technology.

The major change introduced through mobile-related communications technology is that data are collected and processed at large scale in areas that were previously untouched by digitalization. The key question is therefore how the introduction of this technology affects data subjects such as drivers or passengers. What risks are data subjects exposed to; what measures are available to mitigate those risks; and is the unavoidable residual risk acceptable compared to the benefits promised by the technology.

To better understand these questions, it is useful to look at the data that are processed and the insight about peoples' lives they can provide. As probably expected from mobility-related data, they concentrate on time and location. This is for example evident in CAM messages that explicitly broadcast coordinates, but it is also inherent in parking lot reservations that also refer to a place and time.

In data protection, location and time have long been identified as highly critical kinds of data. This has two reasons:

- Location/time data is highly identifying (see for example section 3.3.2 below or section 4.3 "Locations" in [Art29WP 2014] on page 23).
- Location and time enable access to a wide variety of additional data about behavior, preferences, and relations of a person. The following examples shall illustrate this:
    - Frequent location in restricted areas (such as military zones or corporate premises with access control) provides insight in a person's professional occupation and affiliation.
    - Locating persons on hospital premises can provide clues about their health condition.
    - Locating persons repeatedly near places of worship or political activity at the same time of events (such as masses or demonstrations) can reveal religious or political convictions.
    - Repeated co-location of two persons usually indicates a strong relationship; co-location at night time indicates an intimate relationship.

The examples illustrate that the location data can inform about a person's health as well as political and religious convictions. In the GDPR [GDPR 2016], these kinds of data fall under Article 9 and are considered special categories of personal data that warrant particular protection. The location data collected in the context of mobility-related

communications therefore has to be considered to potentially pose a high level of risk for the rights and freedom of natural persons.

In particular, three main risks are present. Namely that legitimate stakeholders or attackers

(i) have access to (location) data that is linked with a long-term identifier of a person,
(ii) can accumulate location data for a given person, and
(iii) can link accumulated data with a long-term identifier of a person.

In risk (i), it is obviously possible to deduct potentially highly sensitive facts about an identified person. The possibility that such information about persons becomes available may for example lead to chilling effects where people don't dare to be seen at certain places or participate in certain activities. Depending on the possible deductions, it may also expose concerned persons to possible discrimination, blackmail, or attacks on reputation.

In risk (ii), the person may not be directly identifiable, but it becomes possible to compile "profiles" or "behavioral patterns" of persons. Since location data possesses a high identification potential on its own, increasing accumulation of data also increases the risk of identification of the actual person. A practical example for this kind of risk is a parking reservation service that uses recurring pseudonymous identifier for a person and can thus link all reservations made belonging to the same person.

In risk (iii), it is now possible to compile profiles of identified persons. This risk can be seen as the amplification of the risks (i) and (ii). In the former case, the accumulation of risk (i) data elements obviously multiplies the risks of a single data element. Actual damage to persons thus becomes more likely and/or more severe. In the latter case, the accumulation of data elements has indeed resulted in the possibility to identify the person. The potential damage is the same as in the former case.

The remainder of this assessment analyzes the presence and magnitude of these risks and possible mitigation measures.

3 CONCEPTS

The analysis of the impact of mobility-related communications technology on data protection requires the use of multiple concepts and terms. In order to base the reasoning on a sound basis, precise definitions are necessary. These are provided in the following.

3.1 Entity

In the context of mobility, an entity is either a (natural) person, a thing (such as a vehicle or a device), a named group of persons (e.g. a family or household), or an organization (or legal person). Obviously, entities can be related to each other in different ways. Other entities, such as animals may be possible, but seem irrelevant in the context of mobility.

3.2 Identity domain

An identity domain is an explicit or implicit management structure for entities. It is closely related to the rules governing the issuance and use of unique identifiers. Many identity domains are managed under the responsibility of an organization or authority.

Identity domains are typically described by properties such as the following:

- **Eligible entities**, i.e., which entities can be managed in a given domain. For example, a national identity domain operated by a government may be limited to persons who are either citizens or residents. A customer identity domain operated by a vendor may be limited to persons or companies who acquire products or services. Some domains may make the conditions for eligibility explicit.
- **Unique Identifier**: Most domains issue a unique identifier such as a social security or a customer number. Some domains may issue multiple identifiers in parallel. For example, due to modernization, several countries currently maintain both, an old and a new social security or tax number.
- **Identifier assignment rules**: These typically control the relationship between entities and assigned identifiers or the change of identifiers over time. Examples include the following:
    - An entity can have only a single identifier.
    - An identifier refers to a single entity
    - An identifier that was assigned to an entity but is no longer in use can or cannot be reassigned to another entity. The relevance of such rules become evident when looking at social security numbers or e-mail addresses.

  Note that assignment rules are often more complex than expected. For example, whiteness protection programs or undercover law enforcement operations may introduce exceptions of the general rules.

3.3 Identifier

An identifier is a data element (e.g., a string or a structured set of data) that is related to an entity. In the following, we also use the term "linked" to an entity. An identifier can be either a unique identifier or a quasi-identifier.

3.3.1 Unique Identifier

A unique identifier is guaranteed to be unique in a given identity domain. In other words, within an identity domain, there is no possibility of "collisions" between identifiers. A unique identifier is liked to a single entity. It is therefore also said, that an entity is uniquely identified by such an identifier.

3.3.2 Quasi-Identifier

A quasi-identifier has a looser relation with an entity. A data element is considered a quasi-identifier, if at least for some eligible entities; it is possible to link the quasi-identifier to an entity with a certain probability. The probability may originate for example in a certain doubt about the linked entity (e.g., and identification with 95% probability) or when the quasi-identifier links to a small set of candidate entities. This definition evidently allows for a wide range of identification strength. Quasi-identifiers are at times unexpected, for example in the case of *religion* and *country* that mostly has a low identification potential, but can in some cases, such as Jews in Iran, be highly identifying (see [Cook 2017]).

The following are examples for common quasi-identifiers:

- **Name, gender, date of birth, place of birth**: As well known from the construction of social security numbers before computers and networks were available, this set of data is highly identifying. Even only the name is a quasi-identifier by itself.
- **Postal Address or location of the home and/or work place**: Even without name and even with limited precision of the location/address, the combination of sleep and work place can in most cases identify an individual person.
- **Postal code, gender, and date of birth:** A study by Latanya Sweeney at Carnegie Mellon University demonstrates how 87% of the American population can be uniquely identified with this data triple [Sweeney 2000].
- **Location and time**: Since entities take up time, knowing the location and time of an entity with good precision can uniquely identify the person. Even with less precision but considering a series of location/time tuples has a very high identification potential. For example, Montjoye et al [Montjoye 2013] demonstrate that in mobile communications, knowing the location only at the precision of the used cell tower and having only a data point per hour, only four spatio-temporal points are enough to uniquely identify 95% of the individuals. Under the title "Anonymization of location data does not work: A large-scale measurement study", Zang and Bolot show that even at large scale, only few

frequent call destinations with a very course location are highly identifying for individuals [Zang 2011].

- **Data derived from location and time**: It has been found that it is very difficult to anonymize location and time related data, for example by deleting location and keeping only speed or acceleration. This is demonstrated for example by Gao et al for speed data [Gao 2014]. Similarly, Devri et al shows how to Inferring Trip Destinations From Driving Habits Data [Devri 2013].
- **Device fingerprints:** Device fingerprints are often highly effective at identifying a device. This has for example be shown for web browsers [Eckersley 2014]. Ferreira Torres and Jonker have studied the possibility of fingerprinting Android apps [Ferreira Torres 2018]. Qiang et al describe all theoretical active and passive fingerprinting possibilities on all layers of a WiFi communications stack [Qiang 2015]. Vo-Huu et al describe a practical study of physical fingerprinting of Wi-Fi devices [Vo-Huu 2016]. They show a success rate of 95+-1% for model identification and 47+-3% for device identification (see table 7, page 12). Baldini et al published a similar study for Vehicle2X communication in C-ITS. They found that in real world conditions, attenuation and fading render the fingerprinting difficult.
- **Unique identifier of a related entity:** The unique identifier of an entity of one kind that is shared by a small group of entities of another kind can be considered a quasi-identifier. For example, this holds for a vehicle identifier of a car that is used only in the family. Evidently, if an entity is used exclusively by another entity, this could even be considered a unique identifier.

3.4 Identification of an entity

In the conceptual framework proposed here, identification of an entity is only possible through the use of an identifier. In particular, there are two kinds of identification:

3.4.1 Deterministic identification

In deterministic identification, a unique identifier is used to determine the single related entity.

3.4.2 Probabilistic identification

In probabilistic identification, a quasi-identifier is used to determine an entity or a small group of entities that are related with the quasi-identifier with certain probability.

3.5 Non-identifying data

When a data element has the potential to identify an entity, it is considered an identifier. Data elements without any potential of identification are called non-identifying data, sometimes also anonymous data or simply data.

3.6 Matching of identifiers

Matching compares two identifiers to determine whether they belong to the same entity. There are two kinds of matching:

3.6.1 Deterministic matching

In deterministic matching, two unique identifiers are compared and considered to belong to the same entity if they are equal. In deterministic matching, only identifiers originating from the same identity domain can be compared.

3.6.2 Probabilistic matching

In probabilistic matching, two quasi-identifiers are compared and considered to belong to the same entity with a certain probability if they are found to be similar or close. The exact semantics of similarity or closeness depend on the type of quasi-identifier used, as illustrated in the following examples:

- **Names**: Data elements representing the name of persons may not always match by equality. Differences in representations of a name may originate from the order of first and last name, the number of first or last names[4] that are captured, possibly truncation of very large names, the use of short or nick names as opposed to the official name (e.g., Bob or Rob instead of Robert), the use of initials, the use of suffixes (such as Senior, Junior, the third), possibly in abbreviated form (Sr., Jr., 3rd), spelling errors, transcoding of special characters such as German Umlauts ('ü' vs. 'ue'), and possibly more. Matching on names is therefore typically based on similarity. The matching is probabilistic, since there may be homonyms[5] and since similar names can only be related with a certain probability.
- **Location**: Location inherently can only be measured with limited resolution and precision. Thus, even measuring the same location twice will likely result in different coordinate values. The matching of location is thus based on proximity, i.e. the distance between two locations.
- **Kinetic Data**: Kinetic data describes the movement of entities at a given point of time. This is for example used in CAM messages where the current location, velocity and acceleration of a vehicle is communicated. Such data allows to estimate the entities' locations in the past and in the future. To match two kinetic data sets, a point of time has to be chosen for which the location is estimated. If the estimate based on one data set is close to that based on the other data set,

[4] Note, that in some countries such as Portugal, last names can consist of a significant number of family names (see for example [Wikipedia Portuguese Name]).

[5] i.e. distinct persons with the same name, see for example http://www.yourdictionary.com/homonym.

there is a probabilistic match. Such matching can be significantly enhanced by taking context into account. For example, matching certainty can be increased by concurrently matching a set of vehicles (and thus accounting for each one), taking into account road maps that restrict assumptions on the possible and probable motion (stay on road, speed limits, corner radius, etc.), or reducing uncertainty by matching on additional data such as vehicle dimensions (that are also contained in CAM messages).

3.7 Linking of distinct data sets

Linking consists of relating two data sets on the basis that they describe the same entity. There are two kinds of linking:

3.7.1 Deterministic linking of distinct data sets

In deterministic linking, two data sets are related since both contain the same unique identifier. In other words, a unique identifier contained in one set deterministically matches with a unique identifier of the other set. From a set theoretical point of view, the sets' intersection contains a unique identifier.

3.7.2 Probabilistic linking of distinct data sets

In probabilistic linking, two data sets are related with a certain probability when they contain a similar or close quasi-identifier. In other words, a quasi-identifier contained in one set probabilistically matches with a quasi-identifier of the other set.

3.8 Intended scope of an identifier

The intended scope of an identifier restricts who can use an identifier for what purpose and for what time period. The following examples shall demonstrate this:

- A transaction pseudonym is used only by the stakeholders who participate in a transaction, only for the purpose of a transaction, and only in the time period necessary to complete the transaction.
- A customer number is typically used only be the customer and the vendor for the purpose of supporting purchases and is valid for the period in which customers are active.

- A nationally unique identification number (such as a social security or tax number) may be used for many different purposes by many stakeholders over the whole life time of the concerned person[6].

There may be various technical measures to enforce the intended scope:

- Removing the identification potential of the identifier (as is typically done for pseudonyms) limits the usefulness of an identifier outside its intended scope and thus removes the motivation of other stakeholders to reuse it for unintended purposes.
- Access control and confidentiality measures can limit the visibility of an identifier to those stakeholders foreseen by the intended scope.
- Deletion can enforce the intended temporal scope.
- Legislation may prohibit the use of an identifier outside its intended scope. This is for example the case in Belgium where the scope of national register number is restricted in article 8 of the law regulating the organization of the national register [Belgium 1983]. Similarly, the GDPR [GDPR 2016], in the opening clause of its article 87, allows member states to restrict the use of national identification number or any other identifier of general application.

3.9 Pseudonym

A pseudonym is an identifier designed to:

1. Prevent the identification of the related entity by unauthorized stakeholders by rendering identification prohibitively difficult or costly.
2. Optionally enable authorized stakeholders only to identify the related entity through the use of additional data (such as a cryptographic key or a lookup table).

In the case that identification is possible, the access to the necessary additional data must be closely controlled and restricted to authorized stakeholders for authorized purposes under well-defined conditions.

3.10 Credential

In communications, it is often not sufficient, that participants themselves claim a certain identity by presenting a unique identifier. It is necessary to provide means that enable participants to trust the identifiers presented by others. Credentials are used to provide trustworthiness to an identifier.

---

[6] For example in Italy, the Codice Fiscale is not only used for tax purposes, but also in every other branch of government (including health care) as well as by banks, insurances or in any formal declaration or contract.

Credentials typically consist of the following parts:

1. A unique identifier of the entity,
2. optionally additional data describing the entity,
3. an identifier of a trusted party that issued the credential,
4. a means for verifying that the credential is valid,
5. a means for verifying that the credential was indeed issued by the claimed trusted party.
6. a means for verifying that it is indeed the entity described by the identifier who presents the credential, and

Examples for credentials are X.509 certificates [Cooper 2008] or SAML bearer assertions [OASIS 2005a]. The different parts of a credential shall be further illustrated using these examples:

In X.509 certificates, the unique identifier is typically part of the *subject common name*, most commonly either the *distinguished name* or the *subject serial number*. The issuer is declared in the *issuer* field. The validity of the certificate is verified by comparing the current date to the *validity* period. In addition, extensions like the *extended key usage* can be used for further validation. The issuer's signature of the certificate allows verifying point 5. It can be verified with an appropriate cryptographic protocol (such as TLS) whether the entity presenting the certificate is its legitimate holder. This is done though inclusion of the entities public key in the certificate (*subject public key information*) and a challenge and response handshake (e.g., the TLS handshake) that proofs possession of the related private key.

In a SAML bearer assertion, the *saml:NameID* within the *saml:Subject* contains a possibly short-lived unique identifier and a longer lived identifier of the entity can be declared in the *uid saml:Attribute*. SAML is perfectly suited to express other data about the entity in different kinds of *saml:Attribues*. For example, the *mail* attribute provides an e-mail address. The issuer of a SAML assertion is declared in the *saml:Issuer*. The validity of the credential can be verified based on the *saml:Condition* element with its *NotBefore* and *NotOnOrAfter* attributes. The issuer is verified through an XML-signature contained in *saml:Signature*. Bearer assertions (as opposed to holder of key assertions) fail to use cryptographic keys to proof that the assertion is presented by the legitimate holder. Instead, there are other (if weaker) means to verify that the assertion is not presented after theft. These verifications for example involve the *AudienceRestriction* element and the verification, that the subject's browser was "directly redirected" (see [OASIS 2005b] section 7.1.1.3, page 24).

4 MODEL OF COMMUNICATIONS

In a mobility setting, transactions can use communications over multiple channels. As a basis for the analysis of linkability in such transactions, this section lays out a model of communications.

4.1 Communication Stacks

Technical communication systems are typically structured in multiple layers with the physical layer at the bottom and the actual application message at the top. This is best known from the OSI Layer Model (see ISO/IEC 7498-1 or [Wikipedia OSI_model]**)** and is depicted in Figure 1.

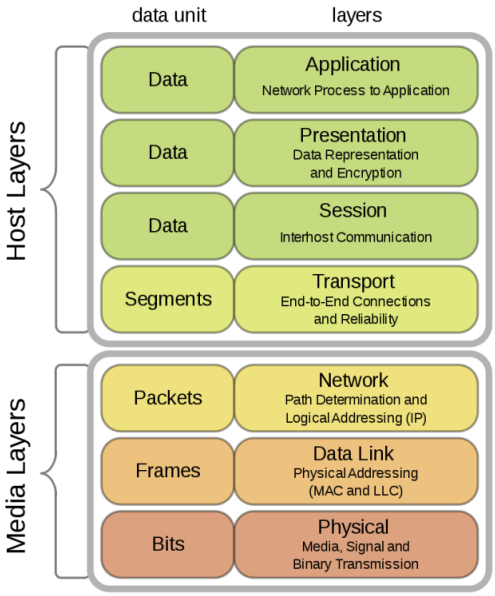

Figure 1: The OSI Model[7]

Other layer models apart from OSI exist. The exact number of layers varies with the model used. This is illustrated by the TCP/IP stack in Figure 2. While the figure leaves away the physical layer, the remaining four layers compare to six layers in the OSI model of Figure 1.

[7] Source: https://commons.wikimedia.org/wiki/File:OSI_Model_v1.svg, File: Osi-model-jb.svg by JB Hewitt Author of SVG edition: Gorivero, GNU Free Documentation License, Version 1.2 and Creative Commons Attribution-Share Alike 3.0 Unported, last visited 7/11/2018.

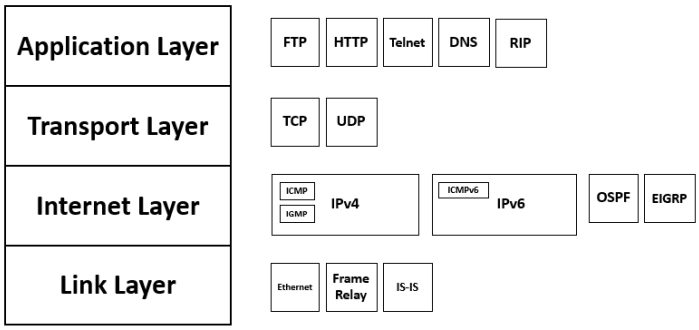

Figure 2: The TCP/IP stack[8]

Figure 3 shows three alternative protocol stacks for HTTP applications. While it operates in the same technology domain as that shown in Figure 2, it illustrates well that the protocols in a stack and their number can vary with the concrete implementation of a communication-based application.

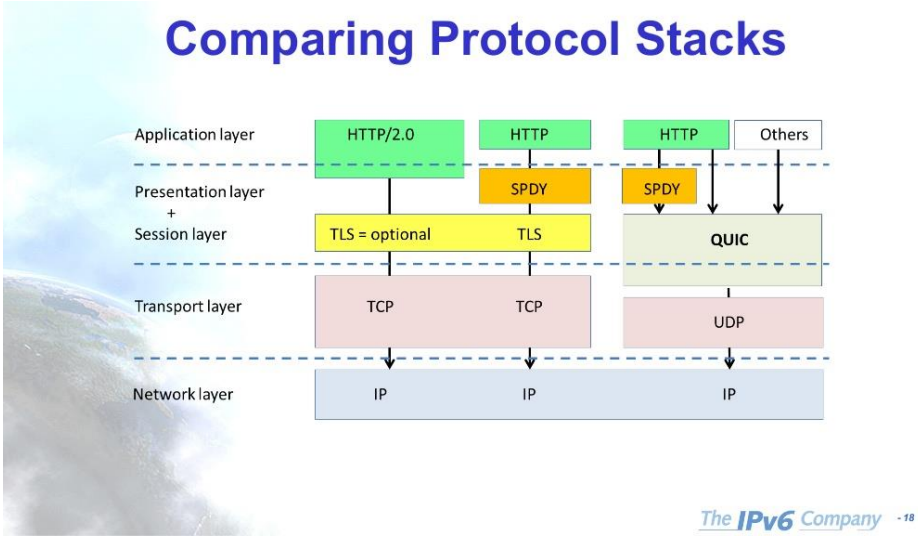

Figure 3: Comparing Protocol Stacks by Jordi Palet[9]

8 Source: https://commons.wikimedia.org/wiki/File:TCP-IP_Model_-_en.png, Author: MichelBakni, Date: 8 October 2017, Creative Commons Attribution-Share Alike 4.0 International, last visited 7/11/2018.

9 Source: Jordi Palet, A new Internet? Intro to HTTP/2, QUIC, DoH and DNS over QUIC, APNIC46 Conference, Noumea, New Caledonia, 6 to 13 September 2018, https://www.slideshare.net/apnic/a-new-internet-intro-to-http2-quic-doh-and-dns-over-quic/18, last visited 7/11/2018.

4.2 Identifiers used in Communications Stacks

Most of the various layers of communications stacks introduce identifiers for different purposes. This section provides some examples for this to introduce the concept. More examples for stacks that are particularly relevant in the context of mobility will be given later.

The physical layer of communications is typically implemented by communications devices. Due to manufacturing tolerances and slight variation between individual devices even of the same make and model, it is often possible to fingerprint the devices.

Fingerprints, while not unique, often permit to distinguish between or in some cases even identify individual devices. Fingerprints can therefore be seen as quasi-identifiers. In the case of personal devices or devices used by a small group of persons, fingerprints can also be used for the probabilistic identification of persons.

Some examples of device fingerprinting from the literature were already provided in section 2.3.2 above.

At the link layer, when devices access a communications media such as a coaxial cable or the airwaves, it is necessary to use a unique identifier that permits to address the device among all other devices connected to the same media. This identifier is called Media Access Control Address or in short MAC Address. Another identifier used by WiFi Access points is the BSSID (see for example [Wikipedia Service set]).

At the network layer, a unique identifier that specifies the destination address of packets is necessary for routing. A prime example for this is the IP Address.

The transport layer may introduce quasi-identifiers. Prime examples are TCP Timestamps. Vanrykel et al illustrate how they can be used to identify the sources of network traffic with a high success rate even when they are hidden behind a network address translation that assigns the same network address to all sources (see [Vanrykel 2017]).

Also the session layer may introduce unique identifiers. This is for example evident when TLS with client certificates are used. Certificates contain many unique identifiers including the certificate serial number, the subject distinguished name, and the public key. In all versions of SSL and in TLS up to version 1.2, the certificates are exchanged in the unencrypted part of the handshake. In TLS 1.3, in contrast, the certificates are encrypted during the exchange.

Also the application layer can introduce a wide variety of identifiers. In this layer reside identifiers such as the following:

- Unique identifiers like user names that are used in the login that is used to restrict access to the application to legitimate users.
- Unique identifiers such as transaction pseudonyms or session IDs that permit to connect multiple accesses in a stateless protocol such as HTTP.
- Pseudo-identifiers such as a browser fingerprint.

Note that there are a variety of mechanisms to communicate identifiers. In a web context it could for example be contained in the HTTP header in the form of a Basic Authentication header or a cookie. Alternatively, it may be conveyed as part of the HTTP data section, for example, as part of a HTML Form.

In the context of communications, it is common to use the terms *data* and *metadata*. Data usually refers to the actual application message that is sent and is restricted to the application layer. But even in the application layer, metadata is exchanged. In a web context, for example, POST data is usually considered data while HTTP headers can already be considered metadata.

The distinction between data and metadata is not relevant for the present discussion of linkability. Instead, all identifiers across all layers of protocol stacks are considered. This section has also illustrated how dangerous it would be to base a study of linkability solely on data while leaving all identifiers of lower layers out of consideration.

4.3 Transaction and Session Identifiers

The previous subsection looked at identifiers used in communication stacks. This usually covers a temporally contiguous exchange between two parties, such as a client and a server. This section looks at the needs of applications to logically group messages over longer periods of time and the kinds of identifiers that this requires.

How the possibilities to logically group messages is limited in the communications model discussed so far is best illustrated at the example of a web service. Here, the client first establishes a TCP connection to the server. This TCP connection represents a session that logically groups all messages sent through it. It is identified by the client's IP address and port number. At a minimum, it groups an HTTP request with the corresponding HTTP response. There may however be many related requests and responses in the same TCP connection. It is never possible, that several applications share the same TCP connection.

A TCP connection has a limited life time. When unused, it times out[10]. Similarly, when the application, such as the browser, that opened the TCP connection is terminated, the TCP connection is also closed.

Many applications require the concept of a session or a transaction that is used to logically group different communication steps.

[10] The time out seems to occur after about two hours, see https://stackoverflow.com/questions/158674/tcp-connection-life (visited on 9/11/2018)

For this purpose, applications issue a specific unique identifier that is used to group logically related communication steps. Typical examples are *session IDs*, *transaction IDs*, or *transaction pseudonyms*. One of many possible ways to manage such identifiers is through the use of HTTP cookies, often called *session cookies*. It is worth noting that cookies are able to span multiple TCP connections.

This subsection has thus introduced an additional type of identifier that is typically used in communications and that will be relevant for the further discussion.

4.4 Actors of Communications

There are usually several actors involved in communications. Figure 4 illustrates those actors that can potentially have access to identifiers that are exchanged in a communication. Actors are represented by rounded boxes and are arranged around a communications stack that is shown in gray. The main actors are shown with full lines, the various attackers are shown with dashed lines.

The probably most obvious actors participating in the communication are the sender and the recipient. In the context of this document, they both operate computing equipment that manages data. This in turn incorporates hardware that is necessary to access the network for sending or receiving data, respectively.

In the case of a direct radio communication between sender and recipient, the communication is conducted “over the air”. In other words, it doesn’t require any network infrastructure that is operated by actors other than the sender and recipient.

In most other cases however, the network is represented by infrastructure that has to be operated. For example, mobile communications require cell towers and a packet network between these towers. The network operator is then usually responsible for operating the towers and the routing of packets between towers.

Larger packet networks, most prominently the Internet, are typically operated by a multitude of network operators. This is for example evident in the fact, that a communication path usually traverses many routers that are owned and operated by different parties.

Apart from the main actors, namely the sender, the recipient, and the network operators, various kinds of attackers may gain access to data. Namely, attackers can target the sender or recipient, respectively, and gain access to data before or after it is communicated. Also, attackers may target the network infrastructure (or the network operators) and obtain data while it is in transfer.

Passive network attackers are usually called “eavesdroppers”. But there are many kinds of active attacks. For example, in “Man in the Middle” (MITM) attacks, the packet routing is manipulated in a way that lets the attacker gain access to and potentially manipulate

the transferred data. In another active attack, the sender could be tricked into communicating with the attacker much rather than with the intended recipient.

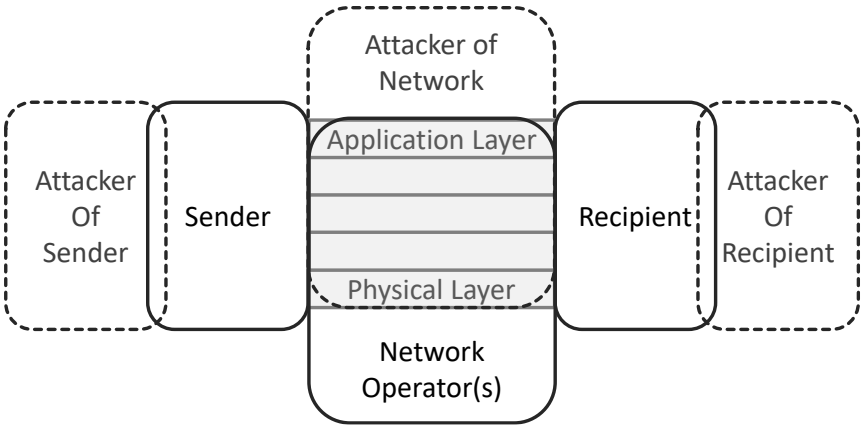

Figure 4: Actors of Communications.

In network communications, different addressing methods are used (see for example [Wikipedia Addressing Methods]). Figure 4 depicts the case of **unicast** where a sender's message is addressed to a single recipient. Also relevant to the discussion at hand are **broadcast**, where a sender addresses all recipients able to receive the message, and **geocast**, where a sender addresses all recipients within a certain range (i.e., distance). Geocast is similar to broadcast in the way the message is sent. There are two main differences to broadcast:

- The recipient verifies to be within the intended range and otherwise drops the message and
- recipients can act as relay stations by rebroadcasting a received message where the range of the initial broadcast is inferior to the range specified in the addressing (see for example [Festag 2014] Figure 3 on page 170). A message can then reach a recipient in multiple hops.

4.5 Visibility of Communications Data by different Actors

In preparation of a discussion of linkability, it becomes relevant which actor is able to see which subset of communications data, most relevantly identifiers. This section briefly describes the need of certain actors to see certain data and the mechanisms that can remove the visibility of certain data. A discussion of concrete cases is left for later.

Accounting for visibility is important to evaluate the potential for misuse of identifiers, for example to identify a person behind a communication, to track that person, or to compile behavioral profiles of that person.

Assuming that in the scenarios that will be discussed later, the sender is usually a person whose privacy needs to be protected; data and identifiers visible to the sender are usually uncritical. The critical actors are therefore the recipient(s), the network operator(s), and the potential attackers. Attackers, depending on their capability, can potentially see all the data that is visible to the actor they attack. While visibility is potentially the same, the difference is that their motivation and willingness for criminal action is largely different from that of the main actors.

While the purpose of the communication is to transfer a message of application data from the sender to the recipient, the above discussion shows that a wide range of identifiers that represent metadata is used in the protocol stacks. Obviously, this data serves a purpose and its visibility to certain actors is necessary in order for them to fulfil their function in the communication.

This is probably most evident for network operators. When they operate a link media, such as a WiFi access point, they obviously need to know link layer identifiers such as MAC addresses for routing packages in their infrastructure. At the higher (inter-)network layer, the recipients network address is obviously necessary for routing of data packets to the recipient.

Similarly, recipients not only need visibility of the application data but also identifiers from lower layers in order to fulfill their function. A prime example is the network address of the sender, without which it would be impossible to send a response to a request.

That said, not all actors need visibility of all identifiers or other data of the stack. For example, consider a heterogeneous (inter-)network where a sender uses connects to a WiFi network to reach a recipient connected over an ISP with a fiber connection. In this case, a message is handled by a multitude of network operators, starting from the WiFi provider over multiple operators of Internet backbone segments to the recipient's ISP. Evidently, identifiers necessary to operate the first WiFi uplink are irrelevant for the tasks of the successive network operators.

An ideal technical design of a communications system would give every actor exactly the visibility that is absolutely needed. Historically however, privacy (and sometimes also security) was not always considered an objective of protocol design. But there is a recent trend of increasingly fixing this situation. Probably the best example is the Internet at large. It started with the transfer of clear text messages that were fully visible to network operators. This was not only the case for the web and HTTP, but also for other kinds of applications such as e-mail (i.e., SMTP). A clear discussion of the trend to add privacy to the Internet was recently given by Jürgen Schmidt ([Schmidt 2018], in German). It describes how an important trigger for the trend were the Snowden revelations, how the web has reached a good degree of privacy protection in the past five years, but that the protection in the rolled out DNS was still badly lacking.

Another example for the trend to restrict identifiers to those necessary for the task was recently reported by Whittaker for the Signal messaging application [Whittaker 2018]. He states: “Dubbed ‘sealed sender,’ the messaging app will soon hide a sender’s information inside the envelope of an encrypted message.” This reveals the sender address to the recipient who can then reply, but prevents access by the network operators and possible attackers.

Another prime example is the redesign of the Transport Layer Security protocol (TLS) [Rescorla 2018]. Du Toit states that “The first TLS working group design goal for TLS 1.3 was to reduce observable data, and they have succeeded for the most part.” (see [Du Toit 2017], page 04, last paragraph of left column). He further writes that “A TLS 1.3 server sends the X.509 certificate, with an optional stapled OCSP response, during the encrypted phase of the handshake” (see same paragraph). This is a major difference to TLS 1.2 where the certificates were exchanged unencrypted. It means that the certificates with their unique identifiers are now invisible to the network operators that need no visibility for fulfilling their task.

Beyond looking at individual protocols and communication stack, the overall design of a possibly complex communicative application should have the goal of limiting visibility of identifiers to what the various actors require for their tasks. Mechanisms to achieve this include the use of suitable protocols, network address translation (many individuals are represented by a single address), the use of (transaction pseudonyms) instead of long-term identifiers, and intermediation designed to hide certain information from other actors. These will be discussed later in this document.

4.6 Context of Single Stack Communications

So far, the present section has considered identifiers and their visibility of a single application on a single communications stack. This subsection explores the wider context of communication to determine whether additional identifiers related to the communication are readily available. For this purpose, it looks at collateral data, i.e., related data that is readily available through services and at other applications of the same sender that may communicate over the same stack.

4.6.1 Collateral Data

Free and for-pay services exist, that allow the lookup of related identifiers when provided with an identifier that is used in a communications stack. This subsection provides examples.

The probably best-known example is the lookup of geolocations from IP addresses. One of the most utilized services is provided by Google [Google 2018] but there is a wide selection of similar services [The ipdata Team 2018]. The lookup of a geolocation from an

IP address was even standardized in the W3C's Geolocation API Specification [Popescu 2016].

It seems that the accuracy of the underlying data bases goes well beyond just a collection of information available from ISPs. [Stackoverflow 2009] and [Superuser 2009] suggest that the data was among others collected by accessing the GPS location and visible WiFi access points from a user's mobile phone while they access a Google service from a given IP address of a PC.

Similar services exist based on the BSSIDs of WiFi access points and for the location of cellular towers of mobile networks.

As reasoned above, geolocations must be considered quasi-identifiers. Using maps, they can be further linked with other quasi-identifiers such as street addresses or ZIP codes or with other data such as "poor residential area" or "restricted military zone".

4.6.2 Other Applications on the Same Communications Channel

Another way for an actor to obtain additional identifiers from the context of communications are the identifiers used by other applications that communicate over the same communications stack.

The following example shall illustrate that. Assume a user is connected to an IP network with dynamic assignment of addresses (i.e., DHCP). Then, for a certain time period, a given user will use the same IP address for different applications run on the same PC.

Further assume that this user runs two applications in this time period. One that is well-designed and avoids any leakage of unnecessary identifiers and a second one that is more leaky, for example by presenting a client certificate in a TLS 1.2 handshake.

In this scenario, a network operator who runs some router along the way of both applications sees solely the user's IP address from the first application, but sees the IP address and the client certificate from the second application.

When only the first application is used at different points of time where different IP addresses are assigned, the network operator cannot link the communications.

This changes when also the second application it used. Now, not only the communications of the second application can be linked on the certificate, but also the communications of the first application that use the same pseudonymous IP addresses as the second application.

This example illustrates how it is important to consider also other applications that communicate over the same stack when determining which identifiers are visible to which actors.

It is interesting to note that a considerable number of applications exist that almost seem to be designed to advertise additional identifiers over the same stack. These will be called

"beacon applications" in the sequel. Such applications have been discovered and studied by Vanrykel et al [Vanrykel 2017]. Like beacons, they leak all kinds of long-term identifiers in clear text over the same channel. Linking such identifiers to well-protected applications without any leaks can be achieved over the IP address, or in case of network address translation that renders this difficult, over the TCP time stamp (that was also discussed above as quasi-identifier).

4.7 Multi-Stack Communications

So far, the identifiers occurring in communications across a single stack and its context have been discussed. This section explores how communications across multiple stack impact on the situation. For this purpose, it looks at the use of multiple stacks for a single transaction and at communications from related stacks that may reveal additional identifiers.

4.7.1 Multiple Stacks used in a Single Transaction

Figure 5 shows a simple scenario of a parking lot reservation service[11]. The single transaction of reserving and then access a parking lot based on this reservation is illustrated. Evidently the transaction consists of multiple steps, each of which using a potentially different communications stack and thus introducing different identifiers.

In the shown scenario, a transaction is initiated by a person still at work who wants to book a parking lot for an after work activity. For this purpose, the person uses a reservation app on a smartphone or tablet to contact a reservation service that knows about the availability of parking lots in different locations. A successful reservation results in a ticket that is managed by the app. This communications is conducted over a normal internet connection using the HTTP protocol.

When arriving at the barrier of the booked parking lot, the car is meant to present the ticket; the latter therefore has to first be transferred from the app to the vehicle. This is done when the person arrives in the car and the smartphone (or tablet automatically) connects with the car's WiFi access point. For this to work automatically, it was necessary to previously couple the smartphone with the access point to enable automatic authorization.

When arrived at the barrier of the parking lot, the vehicle then presents the ticket either over vehicle2x communications or by presenting an RFID.

[11] This scenario is somewhat similar to an iKoPA use case but is modified to keep it much simpler and still show the essence of using multiple stacks in the same transaction.

Before opening the barrier, the ticket has to be validated with the reservation service. For this purpose, an internal IP network is used.

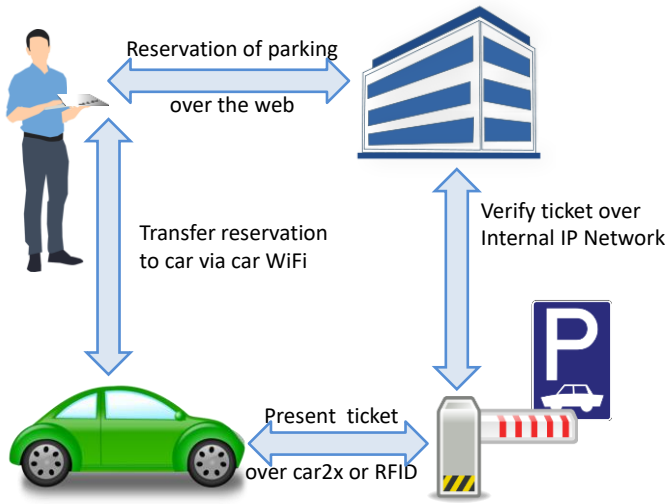

Figure 5: Example of a Transaction involving multiple Stacks.

This example illustrates how real world transactions often involve a multitude of communications stacks. This not only multiplies the identifiers used, but also determines the involved actors. For example, it is evident in the scenario, that a multitude of network operators are involved and that attackers have multiple options to obtain access to various identifiers.

The figure shows a single actor responsible for both, the reservation handling and the parking lot. In a slight but realistic variation, many parking lot operators are likely to collaborate with a few large reservation services. This is way more convenient for drivers who can then cover most of their parking needs with a single reservation app.

This subsection has thus demonstrated that when looking at privacy risks, it is insufficient to look at a single stack with its identifiers, but instead, all involved stacks and actors have to be considered.

4.7.2 Stacks not involved in the Transaction of Interest

The previous subsection has underlined the importance of considering all stacks involved in a transaction of a given application; this subsection reasons that beyond this, other applications and stacks may provide access to additional identifiers of the same person. These therefore need to be identified and taken into account.

In the previous subsection, those stacks were selected that all participate in a common transaction. The “linking” was thus based on transactions. Here, stacks become relevant, since they can be linked based on location and time.

Assume for example that an attacker observes all radio communications of vehicles in front of an entry barrier of a parking lot. Then, not only the presentation of the ticket to the barrier can be observed, but in addition the identifiers of the following other communications stacks:

- CAM messages that are sent every few seconds as part of C-ITS,
- the beacon frames of the car's WiFi access point that can be obtained even through passive scanning (see [Fuxjaeger 2016] page 2, section III A),
- the MAC addresses of the cars' Bluetooth multi-media devices,
- WiFi and Bluetooth MAC Addresses of personal devices carried by drivers or passenger, and
- communications with RFID tags that may be present in the car.

Ullmann et al describe the "secondary vehicle identifiers" based on built-in Bluetooth and WiFi equipment in [Ullmann 2017a]. Fuxjaeger et al describe the observation of MAC Addresses of personal devices in vehicles in motion [Fuxjaeger 2016]. Ali also shows the capturing of Bluetooth addresses from driving vehicles [Ali 2016]. In another paper, Ullmann et al argue that CAM messages can be linked with unique vehicle identifiers of built-in WiFi or Bluetooth equipment (see [Ullmann 2017b], page 41, section VII A).

The wealth of data acquisition equipment available on the market shows how common the detection of WiFi and Bluetooth identifiers in a traffic context already is (see for example [TrafficCast International 2018] and [Traffic Solutions Inc. 2018]).

In addition to these radio-based stacks, license plates can be read with optical sensors even at large scale (see for example [EFF 2018], [Wikipedia ANPR], [Wikipedia ANPR UK], [Parkeon 2018], and [Nikel 2018]).

This discussion shows that a wealth of identifiers is available from other sources than the possibly carefully designed applications that are considered in the foreground of C-ITS. A risk assessment that ignores these possibilities would be incomplete and potentially misleading. The mentioned additional stacks therefore have to be taken into consideration.

5 AVAILABLE MITIGATION MEASURES AND THEIR LIMITATIONS

This section looks at the mitigation measures to address the risks of mobility-related communications technology. It falls short of providing a complete set of measures that consistently results in acceptable residual risks. Much rather, it describes the major currently taken approaches and points out their limitations. The description of such limitations, together with the absence of an obvious solution, may pose questions about how well the privacy of this new technology is under control.

For convenience, the following recalls the risks that need to be mitigated (copied from section 2 above). The three major risks are that legitimate stakeholders or attackers

(i) have access to (location) data that is linked with a long-term identifier of a person,
(ii) can accumulate location data for a given person, and
(iii) can link accumulated data with a long-term identifier of a person.

The various mitigation measures are discussed in the following.

5.1 Pseudonyms

5.1.1 How the risk is mitigated

Risk (i) consists of the possibility to identify the person that is associated with (location) data. This risk has obviously materialized when the location data is associated with a well-known long-term identifier of an entity. The present mitigation measure therefore uses pseudonyms (see definition in section 3.9) in place of long-term identifiers. Re-identification (based on additional data such as keys or lookup tables) shall be either impossible even for data controllers themselves, or rigidly controlled and only possible in well-defined cases under well-defined conditions.

5.1.2 Where the measure is applied

This mitigation measure finds a wide range of application in the field of C-ITS and iKoPA. Most prominently, such pseudonyms are used in authorization certificates that sign CAM messages and to identify reservation transactions in iKoPA (see for example [iKoPA D5v2 2018], section3). The use of pseudonyms was also prominently stated in the iKoPA requirement, as for example in “REQ-PUL-004 Pseudonymization” (see [iKoPA D1v2 2018], page 193).

5.1.3 Effectiveness of the measure

In the areas where this measure is applied, it seems to be effective. This means that undesired re-identification based on the pseudonym is highly unlikely.

5.1.4 Limitations of the measure

The limitations of this measure thus do not lie in a lack of effectiveness, but much rather in the fact that it has not been applied across the board in all areas that require such a measure.

This is most evident in the Bluetooth and WiFi systems present in a vehicle[12]. Whenever they are in use, they leak long-term identifiers such as MAC addresses (see [Ullmann 2017b], page 39, section V B 2). These are not only linkable to the location of observation, but also to locations from CAM messages (see [Ullmann 2017b], page 41, section VII A).

Similarly, even the RFID technology with privacy protection (see [iKoPA D1v2 2018], section 4.3.1.4) uses a long-term Electronic Product Code (EPC) and the privacy features barely limit its visibility. This means that eavesdroppers are now unable to see a recurring identifier of the RFID tag, but parties authorized through a symmetric key[13] will always obtain the same unchanged identifiers. Evidently, an RFID must thus be considered a long-term identifier much rather than a pseudonym.

Besides these radio-based communications stacks, also vehicle license numbers should be mentioned here, since they represent long-term identifiers.

The limitations of this mitigation measure can only be overcome when all linkable long-term identifiers are replaces by pseudonyms. This seems far from easy, however. While legal means are certainly possible[14], technical avoidance of such “identity beacons”, if at all feasible, would be rather complex, costly, and slow to roll out.

It would for example entail the possibility to frequently change WiFi MAC addresses, possibly even during ongoing communications. Changing MAC addresses would however require new versions of the various IEEE 802.11 standards which in turn would then need to be implemented in actual devices and rolled out to consumers.

Similarly, the current RFID standards fail to support pseudonyms and new versions of RFID standards would be required. Also the avoidance of vehicle license plates or the eradication of the currently rather popular automatic license number recognition (ALNR) are difficult to imagine at this point of time.

---

[12] The systems could either be built-in or part of personal devices such as smart phones of passengers.

[13] Such authorized parties are for example parking lot operators whose barriers need to verify reservation tickets presented via RFID.

[14] One could argue that most likely, the GDPR already puts strong limitations on the acquisition of such long-term identifiers. They would have to be required for a legitimate and lawful purpose and it would be necessary to demonstrate that the use of a pseudonym--with much lower risks for the rights and freedom of the concerned data subject—is not a viable alternative to fulfill this purpose.

5.2 Identifier Changes to Limit Possibilities of Accumulation of Data

5.2.1 How the risk is mitigated

Risk (ii) consists in the possibility to accumulate data about a person. Accumulation of data is only possible if it is possible to determine that distinct data sets belong to the same person. As has been reasoned in sections 3.6 and 3.7 above, this is only possible based on the matching of identifiers. The essence of this mitigation measure is therefore to change the (pseudonymous) identifiers in a way to render such matching impossible.

5.2.2 Where the measure is applied

Ideally, this renders aggregation impossible altogether. For example, this was an explicit requirement in iKoPA for parking reservations (see "REQ-PUL-009 Unlinkability of multiple reservations" in [iKoPA D1v2 2018], page 194).

In a weakened form, the identifier changes permit aggregation but limit its possible scale. This seems to be the incentive of changes of pseudonymous authorization certificates in C-ITS (see for example [ETSI 2018b]). Here, one of the possible strategies is to limit the time interval in which a pseudonym is used. The accumulation is then intended to also be limited to this time interval.

5.2.3 Effectiveness of the measure

The following discusses different aspects of effectiveness of this mitigation measure.

5.2.3.1 Effectiveness for different kinds of identifiers

For this discussion, it is essential to note that the considered data (see section 2 above) contains at least two identifiers, a unique pseudonymous identifier and a location as quasi-identifier. For example, a parking reservation contains the transaction pseudonym of the reservation/ticket and the location of the booked parking lot. Similarly, a CAM message contains the pseudonymous authorization certificate and the location coordinates.

For the avoidance of accumulation, it is thus necessary to render matching of all available identifiers impossible. With (pseudonymous) unique identifiers, this is easily achieved through simple change.

With the quasi-identifier of precise kinetic data as they are used in CAM messages, preventing probabilistic matching cannot be achieved with simple change of the location. Here, strategies that prevent matching are more complex. For example, Freudiger et al propose "Mix Zones" [Freudiger 2007]. Here, an interruption of CAM transmissions in a predetermined geographic zone is used, to raise the temporal and locational distance between consecutive data set enough, that probabilistic matching becomes difficult. Evidently, this would not be effective for a lone car on a highway without junctions. It is only then effective, when multiple vehicles are present and "mix" sufficiently, that their future position is not easy to predict.

This example makes it evident, that it is not straight forward to evaluate the effectiveness of strategies to avoid matching of precise location quasi-identifiers. It may for example depend on difficult to control factors as the number of cars that transition a mix zone.

5.2.3.2 Effectiveness for multiple identifiers from a single stack

The discussion so far has focused on the application layer of the stack. Here, in addition, the identifiers from lower levels of the stack are considered.

Evidently, it is mute to change the application layer pseudonym, when lower levels of the stack use the same unique identifier unchanged. Every actor who sees that identifier can then match and consequently accumulate data. The present mitigation measure is thus only then effective, when all unique identifiers throughout the stack are changed in a coordinated fashion.

This is well recognized for vehicle2x messaging. It is for example explicitly addressed by William Whyte in his presentation on the design process and design decisions of IEEE 1609.2[15] (see slide 23 in [Whyte 2016] that is shown in Figure 6). The highlighting of "*Change all identifiers in the stack simultaneously*" was added by the author of this report.

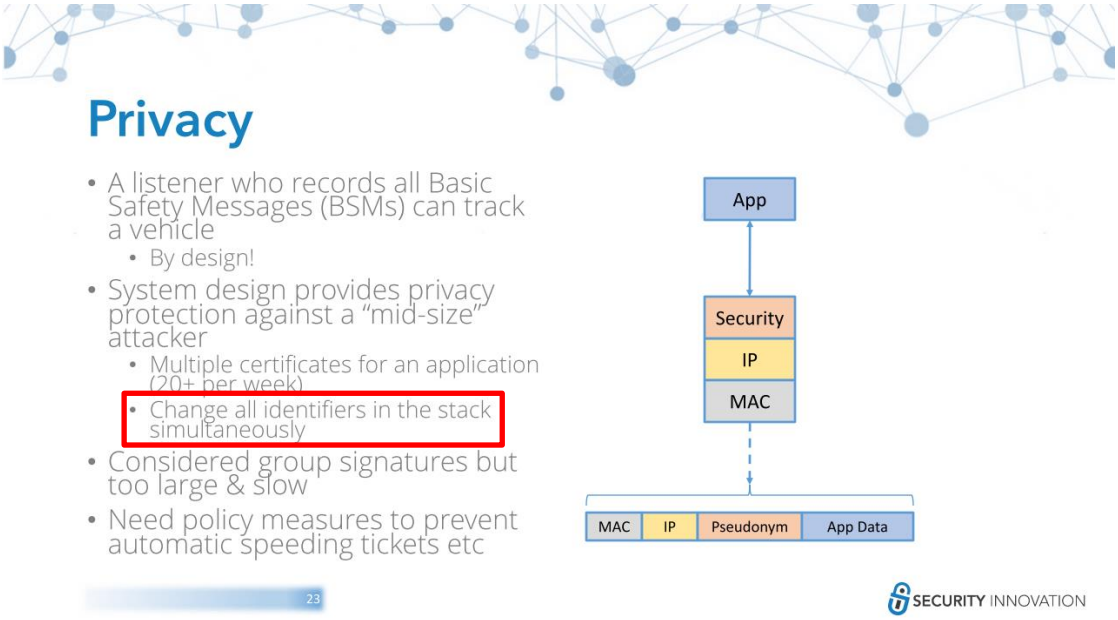

Figure 6: Slide 23 from [Whyte 2016] (red highlighting added by author).

[15] Note that IEEE 1609 is different from C-ITS in as much as it uses application-layer-security instead of network-layer-security. But it was the only explicit statement that the identifiers in the stack are changed simultaneously that the author could find and this surely applies to C-ITS also.

In vehicle2x communications, such as simultaneous change is possible, since the same actor is in control of all layers of the stack.

For other stacks used in mobility-related communications, this may not be the case, however. Consider for example the reservations of parking spaces over a smartphone app. Here, the reservation service obtains the transaction pseudonym presented by the app. Such pseudonyms should be used only once (see section 3.2.4.2 on page 19 in [iKoPA D5v2]). The change-over of such pseudonyms is controlled by the reservation app. On a lower layer of the internet-based communications, the reservation service sees also the IP address. The change of the IP address is controlled by the mobile network operator and may be triggered by the user (e.g., by activating and de-activating flight mode). It may be difficult for an app, however, to trigger a change of IP address.

Note that the unlinkability also depends on the behavior of the provider of the app. In particular, it must refrain from setting any additional long-term identifiers such as cookies or applications reporting their installation identifier (see [Bray 2011]).

This demonstrates that the effectiveness of this measure is not guaranteed for all communications stacks of interest. The effectiveness of a pseudonym change cannot be evaluated limited on the application layer. Instead, a detailed study of the lower layers and the possibilities of changing identifiers *simultaneously* is necessary.

5.2.3.3 Effectiveness for multiple identifiers from multiple stacks

The discussion so far has illustrated that a coordinated change of all identifiers of a single stack may be problematic; this section makes the point that coordinated identifier change is even more difficult when multiple stacks are involved.

This shall be illustrated with a simple scenario that is similar to that of the parking reservation shown in Figure 5 in section 4.7.1 above. Assume that reservations of parking lots and charging stations are made from a smartphone app over the Internet and that the resulting access tickets are presented over vehicle2x to entry barriers or charging stations, respectively.

Figure 7 shows such a scenario in more detail. The Internet stack is shown in light red, the vehicle2x stack in light blue. For every stack, three layers are shown together with their identifiers at different points of time. Assume for simplicity, that the parking provider operates in multiple locations, parking 1 and parking 2 included. Assume further that it operates its own reservation service. Then, the parking operator has visibility of all the identifiers shown in the figure.

From left to right on the time axis, a driver first books parking in two different locations in town from home. After driving to the first location her vehicle presents the reservation ticket at the entry barrier of parking 1. She notices that the vehicle batteries are lower than expected and uses her smartphone to book a charging station to be used later in the second location. Sometime later, when arriving at parking 2, she first presents the parking ticket at the barrier and then the charging ticket at the charging station.

The figure illustrates a situation where in the application layer, every transaction uses a new transaction pseudonym (TP 1 through TP 3) as it should. The authorization layer presents the pseudonym certificates (PC A and PC B) that are used to sign vehicle2x messages. The network (including Media-) layer presents IP addresses in the case of the Internet stack (IP A and IP B) and MAC addresses (MAC a and MAC b) in the case of the vehicle2x stack.

The identifiers are changed periodically. The figure shows identifier change events as triangles on the time axis. It assumes that the change is time-based and that both, the Internet and the vehicle2x stack use the same change frequency. The actual change events are slightly shifted however, since coordination would be rather difficult. Note that the transaction pseudonyms cannot change since they hold together the transactions and changes would result in losing an already made reservation. All other identifiers of the stacks change simultaneously, however, exactly as they should.

The red lines in the figure show the identifier matches that are possible. They make it evident that every communication step can be directly or indirectly linked with every other. In other words, the parking provider is able to link and thus accumulate all this data that refers to the same person.

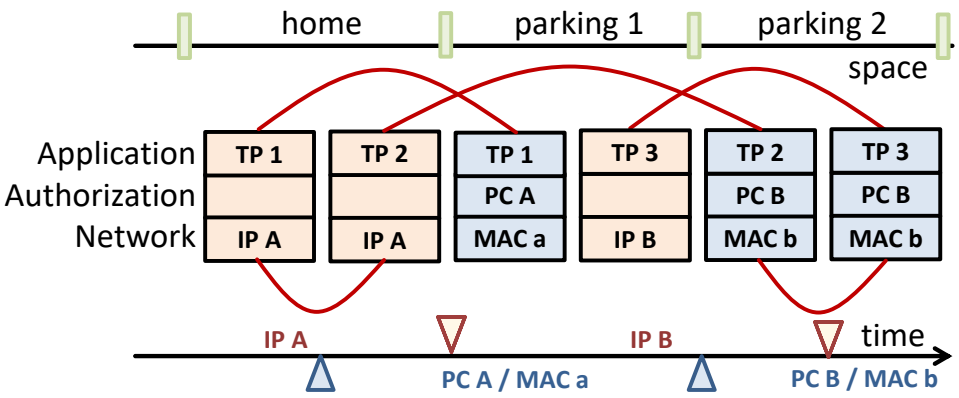

Figure 7: Linking of data due to uncoordinated identifier change.

This scenario illustrates a case that spans multiple communication stacks where unlinkability is not achieved in spite of change of identifiers.

Coordinated change of identifiers is also difficult on stacks (such as the Internet) where multiple applications run concurrently. The transaction of one application may then fall in the middle of a transaction of another application. The change of a network level identifier at the end of the transaction of one application is then likely to affect the proper working of other applications.

Another consideration is that multiple strategies of identifier change exist. The choice of different strategies for different stacks renders coordination of changes even more difficult. The following strategies are in discussion in this context:

- Time-based changes for pseudonymous authorization certificates in the vehicle2x stack (see for example [ETSI 2018b]).
- Location-based as in the Mix Zones also proposed for the vehicle2x stack (see [Freudiger 2007]).
- Transaction-based as is probably the state-of-the-art in privacy-protection of applications (see [Pfitzmann 2008]). This is also used in iKoPA in the reservation service (see section 3.2.4.2 in [iKoPA D5v2 2018]).
- Login-based as currently common in the assignment of IP addresses in mobile networks.

Further work is needed to assess how strongly the difficulty of coordinated identifier change across multiple channels really affects the efficiency of this measure in practice. Also, research on mitigation measures may identify strategies to avoid critical cases altogether.

5.2.4 Limitations of the measure

The limitations of this measure are mostly determined by its effectiveness. While the measure may work in simple settings under control of a single actor, the effectiveness in real world settings with increased complexity is at least questionable. The danger seems to exist that one feels safe because a privacy-oriented measure has been implemented, but that the risk is hardly mitigated and remains high. It is therefore indispensable that the actual effectiveness of this measure are demonstrated in assessments that take the wider context of the application into account.

5.3 Isolation of long-term identifiers from data through the separation of domains

5.3.1 How the risk is mitigated

Risks (i) and (iii) involve the use of long-term identifiers of entities. It is important to understand that long-term identifiers cannot simply be avoided and replaced across the board by pseudonyms. In particular, there are two cases that require long-term identification:

(1) Applications that restrict participation to trusted entities.
(2) Applications that require payment and thus need to reference financial assets of an entity that typically have a long live span.

In the former case (1), the establishment of trust in an entity typically requires data about the entity over a longer period of time. What is needed is an identifier that permits the verification of questions like the following:

- Does the identifier refer to a person or a robot?

- Does the person meet the requirements of eligibility, as for example a minimal age?
- Does the person have a history of misbehavior or has the person accumulated a good reputation?

It is impossible to answer such questions for an entity that is solely identified by a one-time pseudonym. Self-created pseudonyms are therefore unsuited in this case. Much rather, what is needed are pseudonyms issued by authorities or trusted third parties. They know a long-term identifier of the entity, verify eligibility and trustworthiness, and certify this in a (possibly pseudonymous) credential. (See section 3.10 for a definition of credentials).

In the latter case (2), a for-pay service must know who to issue an invoice to. This requires an identifier that refers to financial assets, such as the number of a bank account or credit card. A money transfer from the entity to the service typically reveals this identifier.

In both, risks (i) and (ii), a long-term identifier is linked with (location) data. The mitigation measure discussed here prevents this linking. It achieves this by separating two domains: One where pseudonym is issued based on a long-term identifier and one where the pseudonym is linked to (location) data. The separation works in a sense that

- Stakeholders of the first domain see the full (long-term) identity of the entity and the pseudonym, but lack any access to (location) data.
- Stakeholders of the second domain see the pseudonym and linked (location data), but lack any access to the full (long-term) identity.

One could paraphrase this that in the first domain, the “who” is visible, and in the second domain the “what/where”. A strict separation of these concerns prohibits profiling since the essence of profiling is to link the “who” with the “what/where”.

The following two examples how to separate domains shall illustrate this mitigation measure further. Other approaches, for example based on blind signatures [Chaum 1983] are possible.

What is evident in the examples is that **intermediation** that avoids the direct contact of “who-actors“ and “what/where-actors” is one of the key concepts used to separate domains. Note also the TOR “Onion Router” uses a similar concept of intermediation to prevent profiling [Wikipedia TOR]. Also here, none of the stakeholders (nodes) has access to both, “who” (the client IP address) and “what” (the requested URL).

5.3.1.1 Trusted Pseudonymous Identities

This subsection shows a domain separation that supports case (1) above, i.e., applications that restrict participation to trusted entities. It is illustrated in Figure 8.

The two separate domains are represented by dashed boxes and labeled “identified person” and “pseudonymous identity”, respectively.

In the left domain, a credential issuer knows the full identity of a person, here called the "data subject". In particular, the credential issuer is able to verify the eligibility requirements and if necessary the reputation of the person. On this basis, pseudonymous credentials are issued to the data subject.

The data subjects act as intermediaries between both domains. Since the identity and other data concern the data subjects themselves, the combination of the "who" and the "what/where" are not critical. In this sense, data subject have a very different role from other actors.

In the right domain, data subjects use their pseudonymous credentials to prove their trustworthiness or authorization to participate in the system and make use of offered services. The figure makes it evident that unauthorized entities, such as robots, are unable to access services. Before delivering services, services, i.e. pseudonym consumers, can verify the credentials. One aspect of this is that they accept only credentials from issuers they trust.

The figure makes it further evident that credential issuers cannot profile data subjects since they are isolated from knowing for what purposes the credentials are used. Similarly, pseudonym consumers cannot profile data subjects since they are isolated from their identity and all they see is a short-lived, one-time-use pseudonym.

With the additional actor called "misbehavior resolution", the figure hints at the fact that in exceptional situations, it may be acceptable to break the separations of the domains. It is evident that the conditions when this is desirable must be very well defined and enforced. Any actor who can bridge the two domains must be highly trustworthy and protect the privacy of the data subjects. In the Austrian Citizen Card project, the data protection agency was therefore assigned to this role.

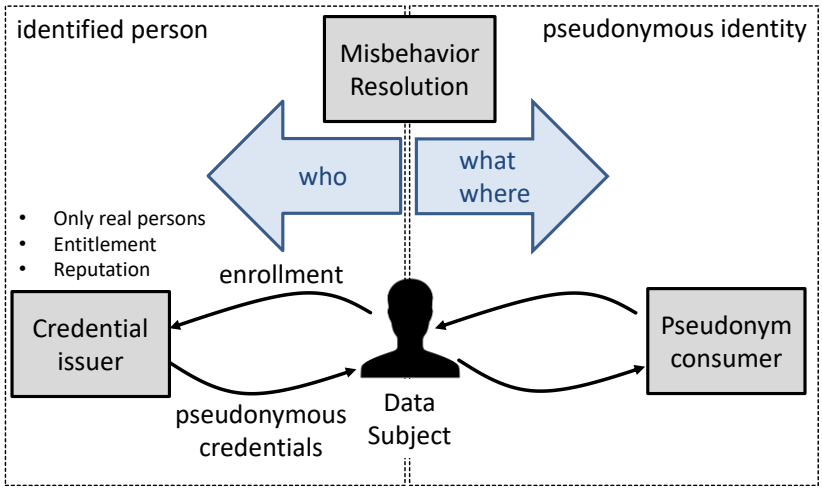

Figure 8: Trusted Pseudonymous Identities.

5.3.1.2 Pseudonymous Payments

This subsection shows a domain separation that supports case (2) above, i.e., applications that require payment and thus need to reference financial assets of an entity. It is illustrated in Figure 9.

The situation is similar to the previous example except that an (at least indirect) interaction between service providers and payment service is necessary on a routinely basis in order to transfer funds that reimburse for delivered services. This contrasts to the previous situation where misbehavior resolution was a rare exception that could be handled by a particularly trustworthy actor.

In the identified domain on the left of the figure, a data subject first enrolls with a payment service. The latter is something like a bank or PayPal where data subjects can open an account that holds funds available for various payments. This account is typically referenced by a long-term identifier. Opening accounts is time-consuming which makes the concept of one-time accounts unpractical.

To give data subjects the possibility to pay in a pseudonymous fashion, the payment service issues what is called “pseudonymous accounts” in the figure. In contrast to other actors, the bank can map these back to their corresponding long-term accounts. The pseudonymous accounts are designed for one-time use and can pay any service up to a given credit limit. Technically, pseudonymous accounts could be implemented by a certificate that states the credit limit and comes with a private key that permits electronic signatures.

In order to be paid for a delivered service, a service provider issues an invoice that states the amount due. The data subject authorizes payment, for example by signing it with the certificate of the pseudonymous account. The signature covers solely the amount section, leaving the recipient of the payment untouched.

In order to receive a transfer of funds, the service provider now sends a batch of authorized payments to a payment broker. The payment broker strips the payment recipients and sends the remaining signed amounts in batches to the according payment services. In response, the payment service transfers the accumulative sum to the payment broker. The latter now redistributes the sum to the payment recipients.

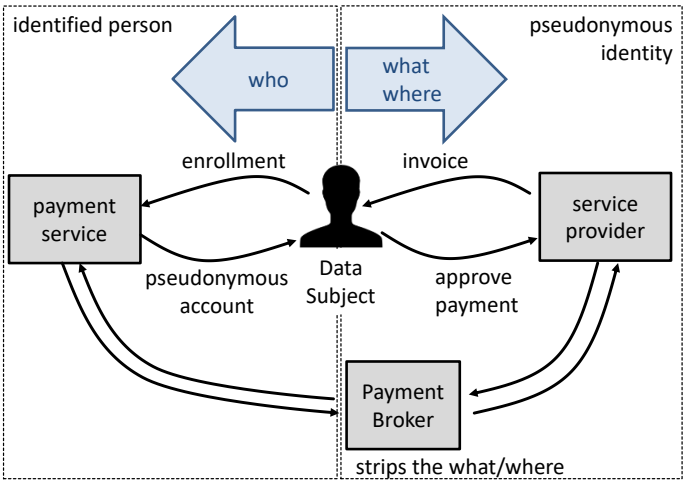

Figure 9: Pseudonymous Payments.

In this setup, none of the actors can profile the data subject: The payment service knows the full identity of the data subject, but what the funds are used for is blinded by the payment broker. The service provider knows what the funds are used for but sees only a one-time pseudonymous identity of the data subject. The payment broker sees the pseudonymous identity and the recipient of funds. The latter may provide some insight in what the funds were used for.

The blinding effect of the payment broker is understood well when comparing to the intermediation by nodes of the TOR onion routing network.

Both of the shown examples were designed to illustrate the concept of domain separation; obviously technically and cryptographically more sophisticated approaches are possible. Among them are privacy Attribute Based Credentials (see for example [ABC4Trust] and David Chaum's eCash [Wikipedia eCash].

5.3.2 Where the measure is applied

The mitigation measure of domain separation is very common. It is for example used in TOR and for anonymous payment systems like eCash. In C-ITS, the authorization certificates that are used in CAM and DENM messages implement this measure.

In iKoPA, the architectural component of "identity provider"[16] was introduced specifically to implement this measure. It acts like the credential issuer in Figure 8 above and issues "certified pseudonyms" (see [iKopa D5v2 2018], section 3.2.4.1).

[16] Note that in previous versions of the architecture, this service was known as "registration service".

5.3.3 Effectiveness of the measure

The measure is usually highly effective in its capability to prevent profiling. There are only few issues that might adversely affect effectiveness.

One is that some actors need to be trusted since could behave in a way that breaks the domain separation. This was already pointed out above about the Misbehavior Resolution authority in Figure 8. It also applies to the Payment Broker in Figure 9 since a failure to strip the information identifying the payment recipient would allow the payment service to compile profiles.

In addition to the trust in individual actor, the effectiveness also requires that actors from different domains don't collude to put their data together. Even if the individual actors can be trusted not to collude, attackers might gain access to data of both domains and link them.

Bearer assertions, i.e. a kind of credential, typically include the intended recipient of the assertion in order to prevent certain attacks (see section 3.10 above). In SAML, the field of the intended recipient is called *AudienceRestriction*. Information about the intended recipient can only be added, if the issuer of the assertion possesses this knowledge. This contradicts the essence of domain separation. Therefore, if bearer assertions are required, an additional intermediary is required for "blinding". The credential issuer then issues all bearer assertions for this intermediary, without having to know about the ultimate recipient. The resulting architecture of a fixed measure is then similar to that of pseudonymous payment in Figure 9.

5.3.4 Limitations of the measure

While the effectiveness of the measure is usually high, the limitation of this measure rather originates from the wider context. Profiling cannot be prevented if in the pseudonymous domain, long-term identifiers of the involved entities can readily be acquired over other channels. This applies for example to the identifiers available from vehicle infotainment systems.

6 SEVERITY OF THE RESIDUAL RISK

In this section, the assessment methodology that was developed for iKoPA is applied to the wider context of mobility-related communications. This results in an initial assessment without ambition of being comprehensive. This section rather attempts to get a first idea of how likely and grave possible attacks are. These attacks are indicators for the residual risk of mobility-related communications after applying the above mitigation measures. This takes into account the limitation of these measures and, even more importantly, that they haven't been applied across the board.

To better understand the situation, the following looks at different factors that influence the residual risk, describes some possible scenarios, and then draws an initial conclusion.

6.1 Attackers and their motivation

The magnitude of risk is significantly determined by the existence of parties who are motivated to obtain data for their purposes (here called attackers). These purposes are typically different from those for which the data was legitimately collected. The data in focus are (location) data of identified persons or accumulations thereof.

A wide group of possible motivations falls under the umbrella of economic value. This fits well with the view that "Data Is the New Oil of the Digital Economy" (see for example [Toonders (2014]). In itself, this means that a large enough collection of data has a value that can be translated into profit. This is certainly a strong motivator.

Concretely, accumulations of identified location data can be used for a multitude of different purposes. They include the following:

- Targeted and location-based advertisement. For example, knowing that a person usually passes a certain service and has just started a trip to that location provides a high success rate to advertisements of this service.
- Traffic management and planning of all sorts is interested in knowing start, destination, and route of trips. Access to background data may be necessary to understand the motivation of the trip or the factors that influence route decisions. Deep learning models may become more accurate and thus competitive on the market, if they can correlate with additional properties of the drivers. Evidently, these kinds of analysis are not possible when pseudonyms change constantly and consequently it is impossible to reconstruct complete routes from start to destination, let alone link to background information that gives insight of possible motivators and makes predications of behavior and thus traffic possible.
- Insurance companies are highly interested to not only understand the risks represented by their customers better. Data sources that are easy to obtain and require less effort than a contractual negotiation with every individual customer might be highly attractive. In disputes with other insurance companies, the data may hold a significant value in determining responsibility.

- Operators of toll roads could be highly motivated to use identified location data. For example, this would basically eliminate the cost of operating toll stations and reduce their adverse effect on traffic flow.
- Employers may be motivated to gain more insight about the travel details of their employees. The sheer possibility that an employer may get to know about private detours may be sufficient to discourage such behavior and save the company possibly significant costs.

Apart from economically motivated uses of the data under discussion, there are also other kinds of motivations. The following examples shall illustrate this.

- The discussed data allows law enforcement to detect violations at a much lower cost than currently possible. This is for example evident when considering the effort that goes in the installation of speed traps. Also the effectiveness of a data-based approach as deterrent is much higher. Instead of only slowing down at the well-known radar trap, drivers are now discouraged anywhere to exceed the speed limit. The large-scale effect of a possible "surveillance" action promises important societal benefits such as fewer lives lost, significant reduction of the number and severity of injuries, protection of school children, and reduction of noise and pollution.
- Jealous spouses may be interested in knowing the whereabouts of their partners.
- Thieves may want to make sure that the inhabitants are far enough from their home and be warned in time should they return.

While this list of attackers and motivation is surely far from comprehensive, it illustrates that there is sufficient incentive to justify a substantial protection.

6.2 Kinds of attacks

The examples above illustrate that there are two kinds of attacks:

- Systematic large-scale attacks where the purposes require large amounts of data about large numbers of data subjects.
- Targeted attacks that are focused on a single data subject.

This distinction is useful when discussing the difficulty to mount certain attacks.

6.3 Cost of mounting an attack

Even with a high motivation of attack, the cost and difficulty of mounting it may significantly affects its likelihood. In particular, the cost may be prohibitively high or reduce the possible scale. Similarly, the difficulty and complexity of mounting an attack may require know-how that is not easily available.

A majority of the critical long-term identifiers that were discussed above are transmitted over radio communications. This renders it accessible to everyone within the transmission range. Compared to wire networks that would have to be attacked with sophisticated means in order to intercept any traffic, this situation represents a very low hurdle.

In addition, ubiquitous off-the shelf equipment can be used to receive these identifiers. For example, Ullmann et al use ordinary notebooks and smartphones as test equipment to receive identifiers over WiFi and Bluetooth (see [Ullmann 2017a], section V. C. “Test Equipment”, on pages 34 and 35).

Vehicle2x communications, i.e., IEEE 802.11p, at the moment still require specialized receivers. But this is likely to change in the short to medium term. ETSI’s ITS architecture already foresees “ITS Personal Stations” (see [ETSI 2010] page 12). The full security promise of the ITS can only be achieved when also pedestrians and cyclists are equipped with ITS stations. The strong societal interest in such protection is for example documented by Schulzki-Haddouti (see [Schulzki-Haddouti 2018], in German). The most promising approach to implementing ITS personal stations is their incorporation in smartphones. These already incorporate WiFi hardware and an extension to support also IEEE 802.11p seems to be a small step. The support of the Vehicle2x channel in smartphones was also predicted by Ullmann et al (see [Ullmann 2017b] page 39). Qualcomm already supports Bluetooth and both, “normal WiFi” (i.e., IEEE 802.11ac) and Vehicle2x (i.e., IEEE 802.11p, aka dedicated short-range communications or DSRC), in a single chipset (namely QCA6584, see [Qualcom 2015], second last paragraph).

This means that the reception of critical identifiers is or shortly will be accessible to everyone who possesses a smartphone.

A possibly remaining hurdle is to position the reception equipment within the range of radio transmission. On first sight, this seems even more challenging when data shall be collected systematically and at large-scale.

Considering that smartphone can be used for the reception of identifiers, this hurdle can be overcome rather easily. All it takes is a popular app or some malware that obtains the identifiers through the in-built WiFi receiver and sends it over the mobile network to a server for collection.

6.4 Risk of detection

The risk of detection of illicit data acquisition with consequent punishment of the responsible parties could drastically lower the likelihood of attacks. This is discussed in the following.

Since passive reception is sufficient to obtain the critical identifiers (see [Fuxjaeger 2016] page 2, section III A), illicit data acquisition is very difficult to detect. A listening station in a parked car, probably using the same antennas as the built in C-ITS equipment, could hardly be recognized even on close inspection.

The same goes for small and well camouflaged “road side” receivers. For example, if they were built into solar lights and placed in gardens near a motorway, they would be impossible to detect.

Considering the scale in which smartphone apps access more data than they possibly need for their functioning and send it to some undeclared servers, an acquisition strategy that abuses the smartphones of normal citizens would also be hard to detect.

But as with Trojans or botnets on PCs, even when their existence is found out and tools for their detection and removal become available, stopping their continued distribution and new infections is very difficult. This particularly since the same surveillance functionality could be easily packaged in different apps and transported by different malware.

Detection could also be a weak deterrent since it would be very difficult to identify the operators of servers placed anywhere in the internet to collect illicit data. Such servers could be operated by legitimate organizations but compromised by hackers.

In summary, the risk of detection is so low that it is hardly a deterrent even for large-scale attacks.

6.5 Attack Scenarios

To further illustrate that the risks are realistic, the following describes two possible scenarios. One is a targeted attack, the other a large-scale one.

6.5.1 Example of a targeted attack

The attack assumes that an attacker wants to follow a targeted individual who drives a car. The prime objective to the attack is to determine the destination of the trip.

In the old days, when this was done for example as part of a police investigation, there were two possible approaches:

- A tracking team followed the targeted vehicle from a safe distance as not to be detected. A regular change of the pursuing vehicle was necessary to further avoid detection. The cost and effort of this tracking method limited its use to a few critical cases.
- A GPS tracking device was secretly fit under the targeted vehicle such that it could be followed with less effort by a single pursuing car that remains within the range of the device’s radio signal. Secret “bugging” a vehicle in this way cannot be done lightly and typically requires a court order to be legal.

In a modern scenario, vehicles can be considered to be bugged by default in the factory. Like a law-enforcement tracking device, the vehicle2x stack sends out CAM messages with coordinates in clear over radio frequencies.

Once vehicle2x communications (IEEE 802.11p) will be supported by smartphone WiFi transceivers, what was formerly possible only for law enforcement in possession of a court

order will become readily accessible to the public. This includes law enforcement officers themselves who likely will no longer require a court order, as well as jealous spouses who distrust their partners.

The change of pseudonyms in the authorization certificates of CAM messages will do little to prevent tracking since the kinetic data will make it straight forward to link to the new pseudonym. Larger mix zones in which no CAM messages are sent may be effective but are predictable and can be overcome for example by classical trailing in a distance that permits visual observation.

6.5.2 Example of a large-scale attack

This attack attempts to compile movement profiles of a large percentage of the population world-wide. As described above, it uses the smartphones of unsuspecting persons to collect the necessary data. The basic idea is that a large enough percentage of the population uses an application or configuration that permits to collect data about vehicles in the surroundings and sends it to a central data collection point.

While this attack may sound far-fetched and difficult to pull off, it is actually happening to date at global scale. The process of the smartphones that collects the data is the Android location service. In the most common configurations, it collects the identifiers of all visible WiFi access points and sends it as input to a central Google service in order to obtain a location in return.

Motivated by the navigation functionality of Google Maps or similar applications, a significant percentage of people driving cars have the location service of their smartphones enabled. Considering that an increasing number of vehicles comes with built-in WiFi access points, the location service obviously reports also the SSIDs/BSSIDs that represent long-term identifiers of the surrounding vehicles to Google.

Since Google compiles and maintains a database of the locations of different WiFi access points[17], it can easily recognize access points that are mobile and thus distinguish between fixed stations and vehicles.

Drivers' smartphones obviously also report the access points of their own vehicle. This makes it possible for Google to link the identifiers of the vehicles' access points to the long-term identifiers of the smartphone. This is a major step towards an identification of the driver. Using the same mechanism, Google can also distinguish between vehicles always driven by the same person and shared vehicles such as rental cars. It could even detect that vehicles are suddenly driven by unexpected persons such as thieves.

[17] Only by knowing the position of access points, the location service can derive locations data on their visibility.

It is important to note that the location service does not only affect persons (and their vehicles) who enable the location service on their phones, but also persons who disable all services that permit tracking but simply use the WiFi access point of their cars.

It is also worth noting that through its location service of the Google controlled Android operating system, Google has established a global monopoly of collecting movement profiles of cars and persons linked to them. In no state under the rule of law would law enforcement agencies by permitted to collect only fractions of this data.

7 CONCLUSIONS

iKoPA technology has been developed by following a process of data protection by design. This has resulted in an architecture that mitigates data protection risks as much as possible with the current state of the art. The residual data protection risk of iKoPA technology has thus been found to be small.

The present document has described the methodology that was developed for the impact assessment of iKoPA technology. This methodology has then been applied beyond iKoPA technology to the wider context of mobility-based communications. This initial study has already identified significant areas with yet unmitigated and at times considerable risks. This confirms the reusability of the developed concepts and methodology.

The impact assessment has focused on the risk of collecting location data of identified persons and the collection of such data into movement profiles. It has analyzed the mitigation measures that are common in mobility-related communications, namely pseudonymization, frequent and coordinated change of pseudonyms, and separation of domains. The analysis included an assessment of the effectiveness and limitations of these measures.

On this basis, the severity of the residual risk of mobility-related communications after mitigation was discussed. It found that a multitude of highly motivated and capable attackers exist, that attacks are relatively simple to deploy and hard to detect, and that even at this point of time, a systematic attack that collects movement data of modern vehicles is under way. It is important to note that this attack does not involve iKoPA technology.

The report therefore concludes that currently, the residual risk in mobility-related communications is very considerable and recommends that urgent actions be taken to address the current shortcomings. The following is an incomplete list of possible actions:

- The biggest danger stems from the current systematic large-scale collection of vehicle location data most likely in a third country. This is made possible by WiFi access points that are built into vehicles and lack any mitigation measure such as pseudonymization, together with the existing smartphone location services by Google and Apple. It is important to make policy makers aware of this issue, that it is studied in more detail, and that options to mitigate the problem are explored.
- The current generation of privacy-enhanced RFIDs still lacks the possibility of assigning changing pseudonymous TAG identifiers. An improved generation of RFIDs needs to be developed before wide-spread application in a mobility setting can be considered.
- The combined use of a multitude of communication technologies and channels even within the same transactions raises new risks that are not yet fully understood. Therefore, additional research is necessary to address this issue (see

below). It is also recommended that data protection working groups of C-ITS also address the wider scope of mobility-related communications.

The initial success also invites future research that applies the iKoPA methodology more systematically and thoroughly. Such research could include the following:

- A detailed study of identifiers of all commonly used communication stacks including the visibility of identifiers to the different actors of the communications.
- On this basis, a number of common use cases and scenarios of mobility-related communications can be assessed.
- To push the state of the art of mitigation measures, a more detailed study could address multi-channel pseudonym change strategies and the likelihood of reduced efficiency of this measure when full coordination across channels is not possible.

8 BIBLIOGRAPHY

ABC4Trust Collaborative project funded by the EC within the 7th Framework Programme, https://www.abc4trust.eu/, last visited 22/11/2018.

Acar (2017) Günes Acar, Online Tracking Technologies and Web Privacy, PhD Thesis, KU Leuven, ARENBERG DOCTORAL SCHOOL, Faculty of Engineering Science, May 2017, https://www.esat.kuleuven.be/cosic/publications/thesis-289.pdf, last visited 09/10/2018.

Ali (2016) Junade Ali, Tracking Drivers with Bluetooth, IcyApril, November 21, 2016, https://icyapril.com/privacy/2016/11/21/tracking-drivers-through-their-phones.html, last visited 10/10/2018.

Art29WP (2014) ARTICLE 29 DATA PROTECTION WORKING PARTY, WP 216, Opinion 05/2014 on Anonymisation Techniques, Adopted on 10 April 2014, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf, last visited 14/11/2018.

Art29WP (2017) ARTICLE 29 DATA PROTECTION WORKING PARTY, WP 252, Opinion 03/2017 on Processing personal data in the context of Cooperative Intelligent Transport Systems (C-ITS), Adopted on 4 October 2017, https://ec.europa.eu/newsroom/just/document.cfm?doc_id=47888, last visited 02/10/2018.

Baldini (2017) Gianmarco Baldini, Raimondo Giuliani, and Eduardo Cano Pons, An Analysis of the Privacy Threat in Vehicular Ad Hoc Networks due to Radio Frequency Fingerprinting, in Mobile Information Systems, Volume 2017, Article ID 3041749, 13 pages, https://doi.org/10.1155/2017/3041749, last visited 10/10/2018.

Belgium (1983) Belgium, Loi organisant un registre national des personnes physiques, August 8, 1983, http://www.ejustice.just.fgov.be/eli/loi/1983/08/08/1984021127/justel, last visited 18/10/2018.

Bray (2011) Tim Bray**,** Identifying App Installations, Android Developers Blog, 30 March 2011, https://android-developers.googleblog.com/2011/03/identifying-app-installations.html, last visited 16/11/2018.

Chaum (1983) David Chaum, Blind signatures for untraceable payments, Advances in Cryptology Proceedings of Crypto. 82 (3): 199–203,

http://www.hit.bme.hu/~buttyan/courses/BMEVIHIM219/2009/Chaum.BlindSigForPayment.1982.PDF, last visited 22/11/2018.

C-ITS Platform DPWG (2017), Processing personal data in the context of C-ITS, 10/07/2017, Document prepared by the Data Protection and Privacy Working Group of the C-ITS Platform for Art. 29, contained in https://ec.europa.eu/transport/sites/transport/files/2017-09-c-its-platform-annexes.zip, filename "Data Protection - Annex I - Processing personal data in the context of C-ITS.docx", last visited on 02/10/2018.

Cook (2017) John D. Cook, Toxic pairs, re-identification, and information theory, 30 September 2017, https://www.johndcook.com/blog/2017/09/30/toxic-pairs/, last visited 18/10/2018.

Cooper (2008) D. Cooper, S. Santesson, S. Farrell, S. Boeyen, R. Housley, and W. Polk: Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile, IETF, Network Working Group, RFC 5280, https://tools.ietf.org/html/rfc5280, last visited 1/11/2018.

Devri (2013) Rinku Dewri, Prasad Annadata, Wisam Eltarjaman, Ramakrishna Thurimella: Inferring Trip Destinations From Driving Habits Data, WPES 2013, http://cs.du.edu/~rdewri/data/MyPapers/Conferences/2013WPES-Extended.pdf, last visited 17/10/2018.

Du Toit (2017) Roelof Du Toit, Responsibly Intercepting TLS and the Impact of TLS 1.3, Symantec Technical Brief, https://www.symantec.com/content/dam/symantec/docs/other-resources/responsibly-intercepting-tls-and-the-impact-of-tls-1.3-en.pdf, last visited 9/11/2018.

EC (2016a) C-ITS Platform, Final report, January 2016, https://ec.europa.eu/transport/sites/transport/files/themes/its/doc/c-its-platform-final-report-january-2016.pdf, last visited 01/10/2018.

EC (2016b) Research Theme Analysis Report, Cooperative Intelligent Transport Systems, http://www.transport-research.info/sites/default/files/brochure/TRIP_C-ITS_Report.pdf, last visited 05/10/2018.

EC (2017) C-ITS Platform, Final report Phase II, September 2017, https://ec.europa.eu/transport/sites/transport/files/2017-09-c-its-platform-final-report.pdf, last visited 02/10/2018.

EC C-ITS Platform, Intelligent transport systems, Cooperative, connected and automated mobility (CCAM), web site, https://ec.europa.eu/transport/themes/its/c-its_en, last visited 02/10/2018.

Eckersley (2014) Peter Eckersley. 2010. How unique is your web browser?. In Proceedings of the 10th international conference on Privacy enhancing

technologies (PETS'10), Mikhail J. Atallah and Nicholas J. Hopper (Eds.). Springer-Verlag, Berlin, Heidelberg, 1-18, https://panopticlick.eff.org/static/browser-uniqueness.pdf, last visited 17/10/2018.

EFF (2018) Electronic Frontier Foundation, Automated License Plate Readers (ALPRs), https://www.eff.org/pages/automated-license-plate-readers-alpr, last visited 15/11/2018.

Eisses et al (2012) Stefan Eisses, Tom van de Ven, Alexandre Fiev, ITS & Personal Data Protection, Final Report, Amsterdam, October 4th, 2012, 20121004_ITS AP5 1_D5 Final Report v1.0 SEI.docx, prepared for the EUROPEAN COMMISSION, Directorate-General Mobility and Transport, FRAMEWORK CONTRACT TREN/G4/FV-2008/475/01 https://ec.europa.eu/transport/sites/transport/files/themes/its/studies/doc/2012-its-and-_personal-data-protection_-_final_report.pdf, last visited 04/10/2018.

Englehardt (2015) Steven Englehardt, Dillon Reisman, Christian Eubank, Peter Zimmerman, Jonathan Mayer, Arvind Narayanan, and Edward W. Felten. 2015. Cookies That Give You Away: The Surveillance Implications of Web Tracking. In *Proceedings of the 24th International Conference on World Wide Web* (WWW '15). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 289-299. DOI: https://doi.org/10.1145/2736277.2741679, https://jonathanmayer.org/publications/www15.pdf, last visited 09/10/2018.

ETSI (2010) ETSI TS 102 637-1, V1.1.1, 2010-09, Technical Specification, Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 1: Functional Requirements, https://www.etsi.org/deliver/etsi_ts/102600_102699/10263701/01.01.01_60/ts_10263701v010101p.pdf, last visited 23/11/2018.

ETSI (2018a) ETSI TS 102 941, V1.2.1, 2018-05, Technical Specification, Intelligent Transport Systems (ITS); Security; Trust and Privacy Management, https://www.etsi.org/deliver/etsi_ts/102900_102999/102941/01.02.01_60/ts_102941v010201p.pdf, last visited, 04/10/2018.

ETSI (2018b) ETSI Technical Report TR 103 415, V1.1.1, 2018-04, Intelligent Transport Systems (ITS); Security; Pre-standardization study on pseudonym change management, https://www.etsi.org/deliver/etsi_tr/103400_103499/103415/01.01.01_60/tr_103415v010101p.pdf, last visited on 02/10/2018.

Ferreira Torres (2018) Ferreira Torres C., Jonker H. (2018) Investigating Fingerprinters and Fingerprinting-Alike Behaviour of Android Applications. In: Lopez J., Zhou J., Soriano M. (eds) Computer Security. ESORICS 2018. Lecture Notes in Computer Science, vol 11099. Springer, Cham,

https://www.researchgate.net/profile/Hugo_Jonker2/publication/326579507_Investigating_Fingerprinters_and_Fingerprinting-alike_Behaviour_of_Android_Applications/links/5bab8431299bf13e604cdf70/Investigating-Fingerprinters-and-Fingerprinting-alike-Behaviour-of-Android-Applications.pdf, last visited 17/10/2018.

Festag (2014) Andreas Festag, Cooperative intelligent transport systems standards in Europe, Communications Magazine, IEEE. 52. 166-172, December 2014, 10.1109/MCOM.2014.6979970, https://www.researchgate.net/publication/273395691_Cooperative_intelligent_transport_systems_standards_in_Europe, last visited 9/11/2018.

Freudiger (2007) J. Freudiger, M. Raya, M. Félegyházi, P. Papadimitratos, and J.-P. Hubaux, Mix-Zones for Location Privacy in Vehicular Networks, In Proceedings of WiN-ITS, August 2007, https://infoscience.epfl.ch/record/109437/files/FreudigerRFPH07winITS.pdf?version=2

Friedewald (2018) Michael Friedewald, Felix Bieker, Hannah Obersteller, Maxi Nebel**,** Nicholas Martin, Martin Rost, Marit Hansen, White Paper, DATENSCHUTZ-FOLGENABSCHÄTZUNG, Ein Werkzeug für einen besseren Datenschutz, Dritte, überarbeitete Auflage, FORUM PRIVATHEIT UND SELBSTBESTIMMTES LEBEN IN DER DIGITALEN WELT, Herausgeber: Michael Friedewald, Regina Ammicht Quinn, Marit Hansen, Jessica Heesen, Thomas Hess, Jörn Lamla,Christian Matt, Alexander Roßnagel, Sabine Trepte, Michael Waidner, https://www.forum-privatheit.de/forum-privatheit-de/publikationen-und-downloads/veroeffentlichungen-des-forums/themenpapiere-white-paper/Forum-Privatheit-WP-DSFA-3-Auflage-2017-11-29.pdf, last visited 16/10/2018.

Fuxjaeger (2016) Paul Fuxjaeger, Stefan Ruehrup, Thomas Paulin, and Bernd Rainer: Towards Privacy-Preserving Wi-Fi Monitoring for Road Traffic Analysis, IEEE Intelligent Transportation Systems Magazine, 8, 63-74, 10.1109/MITS.2016.2573341, https://www.researchgate.net/publication/305877717_Towards_Privacy-Preserving_Wi-Fi_Monitoring_for_Road_Traffic_Analysis, last visited 13/11/2018.

Gao (2014) Xianyi Gao, Bernhard Firner, Shridatt Sugrim, Victor Kaiser-Pendergrast, Yulong Yang, Janne Lindqvist, Elastic Pathing: Your Speed is Enough to Track You, UbiComp 2014, https://www.winlab.rutgers.edu/~janne/elasticpathing-ubicomp14.pdf, last visited 17/10/2018.

GDPR (2016), REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), European

Legislation Identifier http://data.europa.eu/eli/reg/2016/679/oj, last visited 16/10/2018.

Google (2018) Geolocation API, Developer Guide, https://developers.google.com/maps/documentation/geolocation/intro, last visited 9/11/2018.

Hansen (2015) Hansen, Marit / Jensen, Meiko / Rost, Martin, 2015: Protection Goals for Privacy Engineering, Proceedings for the International Workshop on Privacy Engineering, IWPE'15.

iKoPA D1v2 (2018) Deliverable D1v2: Requirements Analysis and System Architecture, Version 2.0, https://ikopa.de/en/results/.

iKoPA D4v2 (2018) Deliverable D4v2, Assessment and recommendations, Version 2.0, https://ikopa.de/en/results/.

iKoPA D5v2 (2018) Deliverable D5v2: Security Requirements and Architecture, Version 2.0, https://ikopa.de/en/results/.

iKoPA Website (2018) Integrated cooperating platform for automated electric vehicles, https:// https://ikopa.de/en/home/, last visited 01/10/2018.

IWGDPT (2018) International Working Group on Data Protection in Telecommunications, Working Paper: Connected Vehicles, Budapest, Hungary, April 9/10, 2018, https://www.datenschutz-berlin.de/fileadmin/user_upload/pdf/publikationen/working-paper/2018/2018-IWGDPT-Working_Paper_Connected_Vehicles.pdf, last visited 16/10/2018.

Montjoye (2013) Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen & Vincent D. Blondel, Unique in the Crowd: The privacy bounds of human mobility, Nature, Scientific Reports 3, Article number: 1376 (2013), http://rdcu.be/JGxh, last visited 17/10/2018.

Nikel (2018) David Nikel, Road Tolls in Norway, https://www.lifeinnorway.net/road-tolls-in-norway/, last visited 10/10/2018.

OASIS (2005a) Scott Cantor, John Kemp, Rob Philpott, Eve Maler (eds.), Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0, OASIS Standard, 15 March 2005, Document identifier: saml-core-2.0-os, Location: http://docs.oasis-open.org/security/saml/v2.0/, http://docs.oasis-open.org/security/saml/v2.0/saml-core-2.0-os.pdf, last visited 1/11/2018.

OASIS (2005b) Frederick Hirsch, Rob Philpott, Eve Maler (eds.), Security and Privacy Considerations for the OASIS Security Assertion Markup Language (SAML) V2.0, OASIS Standard, 15 March 2005, Document identifier: saml-sec-consider-2.0-os, Location: http://docs.oasis-open.org/security/saml/v2.0/, https://docs.oasis-

open.org/security/saml/v2.0/saml-sec-consider-2.0-os.pdf, last visited 05/11/2018.

Parkeon (2018) Parkeon Ldt., MiniPark ANPR Parking System, https://www.parkeon.co.uk/our-solutions/product-catalogue/anpr-parking-system/, last visited 10/10/2018.

Pfitzmann (2008) Andreas Pfitzmann and Marit Hansen, Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management, Version v0.34, 10 August 2010, http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf, last visited 16/11/2018.

Popescu (2016) Andrei Popescu (ed.), Geolocation API Specification 2nd Edition, W3C Recommendation 8 November 2016, https://www.w3.org/TR/geolocation-API/, last visited 9/11/2018.

Qiang (2015) Xu, Qiang & Zheng, Rong & Saad, Walid & Han, Zhu. (2015). Device Fingerprinting in Wireless Networks: Challenges and Opportunities. IEEE Communications Surveys & Tutorials. 18. 10.1109/COMST.2015.2476338. https://www.researchgate.net/publication/270593911_Device_Fingerprinting_in_Wireless_Networks_Challenges_and_Opportunities, last visited 10/10/2018.

Qualcomm (2015) Qualcomm Drives Future of Automotive Connectivity with New 4G LTE Modems, https://www.qualcomm.com/news/releases/2015/03/02/qualcomm-drives-future-automotive-connectivity-new-4g-lte-modems, last visited 06/12/2018.

Rescorla (2018) E. Rescorla, The Transport Layer Security (TLS) Protocol Version 1.3, IETF, Network Working Group, RFC 8446, https://tools.ietf.org/html/rfc8446, last visited 9/11/2018.

Rost (2002) Rost, Martin, 2012: Standardisierte Datenschutzmodellierung; in: DuD - Datenschutz und Datensicherheit, 36. Jahrgang, Heft 6: 433-438.

Rost (2009) Rost, Martin; Pfitzmann, Andreas, 2009: Datenschutz-Schutzziele - revisited; in: DuD - Datenschutz und Datensicherheit, 33. Jahrgang, Heft 6, Juli 2009: 353-358.

Rost (2011) Rost, Martin; Bock, Kirsten, 2011: Privacy By Design und die Neuen Schutzziele - Grundsätze, Ziele und Anforderungen; in: DuD - Datenschutz und Datensicherheit, 35. Jahrgang, Heft 1: 30-35.

Rost (2017) Rost, Martin, 2017: Organisationen grundrechtskonform mit dem Standard-Datenschutzmodell gestalten; in: Sowa, Aleksandra (Hrsg.), 2017: IT-Prüfung,

Sicherheitsaudit und Datenschutzmodell, neue Ansätze für die IT-Revision, Wiesbaden, Springer Vieweg: 23-56.

Schmidt (2018) Jürgen Schmidt, Kommentar zu DNS over HTTPS: Die Gruft DNS gehört ausgelüftet, heise online, Forum Sicherheit, 26.10.2018, https://www.heise.de/security/meldung/Kommentar-zu-DNS-over-HTTPS-Die-Gruft-DNS-gehoert-ausgelueftet-4203225.html, last visited 9/11/2018.

Schulzki-Haddouti (2018) Christiane Schulzki-Haddouti, "Schutzranzen"-Projekt kombiniert Kinder-Tracking mit Verkehrssicherheit, 22.01.2018, heise online, https://www.heise.de/newsticker/meldung/Schutzranzen-Projekt-kombiniert-Kinder-Tracking-mit-Verkehrssicherheit-3947907.html, last visited 23/11/2018.

SDM-de (2018)) Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder, AK Technik, Das Standard-Datenschutzmodell, Eine Methode zur Datenschutzberatung und –prüfung auf der Basis einheitlicher Gewährleistungsziele, V.1.1, Erprobungsfassung , von der 95. Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder am 25./26. April 2018 in Düsseldorf einstimmig beschlossen, https://www.datenschutz-mv.de/static/DS/Dateien/Datenschutzmodell/SDM-Methode_V_1_1.pdf, last visited 16/10/2018.

SDM-en (2016) Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder, AK Technik, The Standard Data Protection Model, A concept for inspection and consultation on the basis of unified protection goals, V.1.0 – Trial version, Unanimously and affirmatively acknowledged (under abstention of Bavaria) by the 92nd Conference of the Independent Data Protection Authorities of the Bund and the Länder in Kühlungsborn on 9-10 November 2016, https://www.datenschutz-mv.de/static/DS/Dateien/Datenschutzmodell/SDM-Methodology_V1_EN1.pdf, last visited 16/10/2018.

Silbersack (2005) Michael James Silbersack, Improving TCP/IP security through randomization without sacrificing interoperability, Proceedings, Fourth European BSD Conference, October 22, 2005, http://www.silby.com/eurobsdcon05/eurobsdcon_silbersack.pdf, last visited 10/10/2018.

Smats Traffic Solutions Inc. (2018) Smats Traffic Solutions Inc., TrafficBox, https://www.smatstraffic.com/products/trafficbox/, last visited 10/10/2018.

Stackoverflow (2009) How does Google calculate my location on a desktop? Question posted on 3/11/2009 by Shadi Almosri, Various responders, Stackoverflow, https://stackoverflow.com/questions/1668304/how-does-google-calculate-my-location-on-a-desktop, last visited 9/11/2018.

Superuser (2009) How does Google My Location work? Question posted on 24/7/2009 by Stefano Borini, Various responders, Superuser,

https://superuser.com/questions/12495/how-does-google-my-location-work, last visited 9/11/2018.

Sweeney (2000) L. Sweeney, Simple Demographics Often Identify People Uniquely, Carnegy Mellon University, Data Privacy Working Paper 3, Pittsburgh, 2000, https://dataprivacylab.org/projects/identifiability/paper1.pdf, last visited 17/10/2018.

The ipdata Team (2018) The ipdata Team, What is the Best IP Geolocation API?, Medium, https://medium.com/@ipdata_co/what-is-the-best-commercial-ip-geolocation-api-d8195cda7027, last visited 9/11/2018.

Toonders (2014) Joris Toonders, Yonego, Data Is the New Oil of the Digital Economy, wired, July 2014, https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/, last visited 23/11/2018.

TrafficCast International (2018) TrafficCast International, BlueTOAD Spectra RSU, http://www.trafficcast.com/spectrarsu/index.html, last visited 10/10/2018.

Ullmann (2017a) Markus Ullmann, Tobias Franz, and Gerd Nolden, Vehicle Identification Based on Secondary Vehicle Identifier -- Analysis, and Measurements, in Ullmann, El-Khatib, Vladeanu, and Tsukada (eds.), Proceedings, VEHICULAR 2017, The Sixth International Conference on Advances in Vehicular Systems, Technologies and Applications, Nice, France, July 23 to 27, 2017, pages 32 to 37, https://www.thinkmind.org/index.php?view=article&articleid=vehicular_2017_2_30_30045, last visited 10/10/2018.

Ullmann (2017b) Markus Ullmann, Thomas Strubbe, and Christian Wieschebrink, Misuse Capabilities of the V2V Communication to Harm the Privacy of Vehicles and Drivers, in International Journal on Advances in Networks and Services, Volume 10, Numbers 1&2, 2017, pages 35-43, https://www.iariajournals.org/networks_and_services/netser_v10_n12_2017_paged.pdf, last visited 10/10/2018.

Vanrykel (2017) Vanrykel E., Acar G., Herrmann M., Diaz C. (2017) Leaky Birds: Exploiting Mobile Application Traffic for Surveillance. In: Grossklags J., Preneel B. (eds) Financial Cryptography and Data Security. FC 2016. Lecture Notes in Computer Science, vol 9603. Springer, Berlin, Heidelberg, DOI https://doi.org/10.1007/978-3-662-54970-4_22, contained in [Acar 2017] on pages 151 – 175 or https://pdfs.semanticscholar.org/c723/18ecc5fdfd309deabb8dc23a097637c4b30c.pdf, last visited 09/10/2018.

Vo-Huu (2016) Tien Dang Vo-Huu, Triet Dang Vo-Huu, and Guevara Noubir. 2016. Fingerprinting Wi-Fi Devices Using Software Defined Radios. In Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks (WiSec '16). ACM, New York, NY, USA, 3-14. DOI:

https://doi.org/10.1145/2939918.2939936, https://dl.acm.org/citation.cfm?id=2939936, last visited 10/10/2018.

Wikipedia Addressing methods Anycast, web page, section 1: Addressing methods, https://en.wikipedia.org/wiki/Anycast#Addressing_methods, last visited 9/11/2018.

Wikipedia ANPR Automatic number-plate recognition, Usage, https://en.wikipedia.org/wiki/Automatic_number-plate_recognition#Usage, last visited 15/11/2018.

Wikipedia ANPR UK Automatic number plate recognition in the United Kingdom, https://en.wikipedia.org/wiki/Automatic_number_plate_recognition_in_the_United_Kingdom, last visited 15/11/2018.

Wikipedia eCash Ecash, web page, https://en.wikipedia.org/wiki/Ecash, last visited 22/11/2018.

Wikipedia IEEE_802.11p, IEEE 802.11p, web page, https://en.wikipedia.org/wiki/IEEE_802.11p, last accessed on 02/10/2018.

Wikipedia Portuguese Name, Portuguese name, web page, https://en.wikipedia.org/wiki/Portuguese_name, last accessed on 18/10/2018.

Wikipedia Service set Service set (802.11 network), web page, https://en.wikipedia.org/wiki/Service_set_(802.11_network), last visited 7/11/2018.

Wikipedia TOR Tor (anonymity network), https://en.wikipedia.org/wiki/Tor_(anonymity_network), last visited 22/11/2018.

Whittaker (2018) Zack Whittaker, Signal rolls out a new privacy feature making it tougher to know a sender's identity, techcrunch, 29/10/2018, https://techcrunch.com/2018/10/29/signal-sealed-sender-feature-messaging-security/, last visited 9/11/2018.

Whyte (2016) William Whyte, IEEE 1609.2 and Connected Vehicle Security: Standards Making in a Pocket Universe, Security Standardization Research Workshop December 6, 2016, https://www.researchgate.net/publication/311452790_IEEE_16092_and_Connected_Vehicle_Security_Standards_Making_in_a_Pocket_Universe, last viewed 05/10/2018.

Wikipedia OSI_model, Open Systems Interconnection model (OSI model), https://en.wikipedia.org/wiki/OSI_model, last visited 05/10/2018.

Zang (2011) Zang, Bolot: Anonymization of Location Data Does Not Work: A Large-Scale Measurement Study, MobiCom 2011, https://www.researchgate.net/publication/220926571_Anonymization_of_locati

on_data_does_not_work_A_large-scale_measurement_study, last visited 17/10/2018.