How to Protect Privacy in Floating Car Data Systems

Stefan Rass
Institute of Applied Informatics
System Security Group
Klagenfurt University, Austria
stefan.rass@uni-klu.ac.at

Simone Fuchs
Institute of Smart
System-Technologies
System Security Group
Klagenfurt University, Austria
simone.fuchs@uni-klu.ac.at

Martin Schaffer
Institute of Applied Informatics
System Security Group
Klagenfurt University, Austria
m.schaffer@syssec.at

Kyandoghere Kyamakya
Institute of Smart System-Technologies
System Security Group
Klagenfurt University, Austria
kyandoghere.kyamakya@uni-klu.ac.at

ABSTRACT

Floating Car Data (FCD) is a valuable source of up-to-date traffic information, with a wide range of applications. Active floating car data techniques require drivers to have their vehicles equipped with on-board units regularly transmitting position and velocity information to a central server. Many potential participants are hence reluctant to join FCD projects because of violations of their privacy due to permanent traceability or possible liability in case of speed limit violations. We present a general method for anonymization of floating car data by deriving pseudonyms for trips and samples with the optional ability of relating samples to trips and trips to each other, whilst hiding the identity of a driver, hence protecting his privacy. The resulting concepts are easy to implement and can be used as building blocks for any FCD system with stringent security constraints. The main advantage of our approach is the guaranteed uniqueness of pseudonyms that can be achieved efficiently, i.e. without any communication between vehicles.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems—*Human information processing*; H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Data sharing*; E.3 [**Data Encryption**]: Public key cryptosystems; F.2.1 [**Analysis of Algorithms and Problem Complexity**]: Numerical Algorithms and Problems—*Computations in finite fields*

General Terms

Security, Human Factors

VANET'08, September 15, 2008, San Francisco, California, USA.
Copyright 2008 ACM 978-1-60558-191-0/08/09 ...$5.00.

Keywords

Pseudonyms, Privacy, Anonymization, Discrete Logarithm Problem

1. INTRODUCTION

Floating Car Data (FCD) techniques that use vehicles as active sensors, transmitting their speed and geographic position within a traffic flow, are a well-researched technique for collecting traffic state information. Projects have been conducted during the last decade in various European capitals like Vienna, Paris or Berlin. While the technical challenges of FCD have long been mastered, privacy and security issues of distributing vehicle's positional data are still neglected or ignored. This paper first explains the need for making FCD anonymous by giving examples for privacy issues. Afterwards, a new approach is introduced that protects the privacy of participating drivers and their data against misuse. While most present-day FCD projects get their data from fleet management companies, where privacy is a minor issue, a new and valuable data source is slowly becoming available: more and more private vehicles are equipped with GPS-units for various purposes, like navigation, automatic emergency systems or electronic toll collection. Including these private data into FCD systems would substantially increase the number of available samples. However, while most drivers value up-to-date traffic state information and would gladly contribute their data to a traffic state image, they are wary of giving away their FCD information, because of potential misuse. Making the transmitted data anonymous using sophisticated encryption methods could ensure that neither the driver, nor the data management company or a third party is able to misuse or distort the data to their own advantage.

Within a FCD system, every transmitting device, the so-called onboard-unit (OBU), usually has a unique identifier – the OBU-id. The OBU-id is transmitted with every position sample, which has a unique identifier too (sample-id). Several samples are further aggregated to a common trip and assigned a common identifier, called the trip-id (cf. Section 2). Our approach provides a method to assign samples to a common trip, without ever revealing the trip-id or the OBU-id, thus protecting the driver's privacy. Furthermore, the approach allows for detection of faked samples that may be inserted into the system by an adversary, as well as it prevents existing samples from being reassigned to different trips. Some examples will emphasize the importance of this problem.

What drivers fear most is that law enforcement gets its hands on their floating car data somehow, derives the driven speed/distance ratio and consequently punishes the driver for speeding. Anonymization of the OBU-id guarantees that the driver's identity can never be revealed from the sampled FCD. So, even if the data gets into the wrong hands, drivers' privacy is technically protected by the computational infeasibility of the underlying problems.

Some insurance companies offer special rates for customers with a low mileage per year. The number of driven kilometers is monitored using a GPS-unit. The insurance company must be able to uniquely assign the sample-ids to trips in order to calculate its overall length and then to assign the trip to a designated OBU-id's mileage. Anonymization of the OBU-id prevents third parties from breaking into the system and to assign samples to a wrong OBU-id, thus artificially in-/decreasing a driver's annual mileage. Also, the gathered data can be passed on to traffic management centers for use as FCD, without having to worry about privacy issues, because it is not efficiently possible to compute the OBU-id from the samples or trips.

Likewise, operators of GPS-based electronic toll collection systems must be able to assign a sample to a trip and further on to a designated OBU-id for correctly accounting the road charge, while at the same time preventing third parties from tampering with the samples in order to falsify the number of toll kilometers and decrease the toll amount. Toll collection data is also an additional valuable data source for floating car data systems.

A further possible scenario is the use of falsified samples by a third party for fraudulent misrepresentation of a traffic state image. Currently, free street segments could be shown as congested to other drivers, thus, for instance, ensuring the impostor a strategic advantage for the own fleet management over potential competitors, or other personal advantage over others.

The examples show that there is a variety of potential scenarios for fraud and misuse with floating car data. Drivers and privacy advocates are correct to be suspicious in giving away a vehicle's positional data. The anonymization approach presented in Section 3 shows how floating car data can be acquired, while protecting the driver's privacy and data against misuse at the same time.

Most closely related to our work is [6], in which a blind signature scheme is proposed for anonymous and authenticated floating car data transmission. Contrary to this work, we mainly focus on mechanisms to anonymize trips and samples using changing pseudonyms. The problem of authenticated peers in the FCD transmission chain is not addressed here, and we refer the reader to [6] and references therein. Apart from this, the concept of changing pseudonyms has been widely used for anonymous communication between peers [16, 2, 22], but to the best of our knowledge, none of these has ever been used to protect privacy in floating car data applications [15, 21, 13].

Protection of anonymity of a drivers information imposes a variety of requirements, which we briefly review in the following section, along with components for their realization.

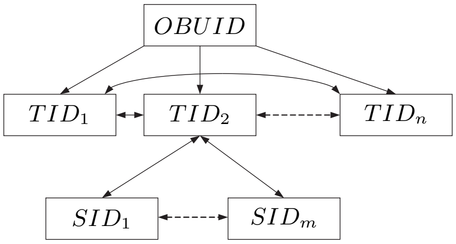

Figure 1: Relationships between different IDs.

2. REQUIREMENTS AND COMPONENTS

In order to preserve privacy for a driver's FCD samples, we adopt a two-stage approach for anonymization of the data. Given a system-wide unique OBU identifier $OBUID$, a straightforward way of formatting FCD sample packets is sending a message of the form $M = (POS, OBUID, t, v)$, where POS is the GPS-position, t is a timestamp, and v is the current velocity. As argued in the previous section, the $OBUID$ should never show up in the sample to prevent from potential data misuse. Neither will a driver want to have his trips to be recorded upon a bunch of FCD samples. A problem for the FCD data collector, however, arises from trips that are interrupted and continue frequently (short stops for lunch, naps, etc.). Further, assume that a new trip-id is chosen each time a vehicle is started after a longer stop (say anything above 15 minutes), such that we spare the driver having to manually invoke a new trip or specifying a short break as no true interruption of a trip. Including the $OBUID$ and the time-stamp can easily fix this issue, as samples can be related to each other trivially, while on the other hand violating a driver's privacy.

Our proposed remedy is using a pseudonym for a trip (trip identifier TID) and several derived pseudonyms for each sample (sample identifier SID). We explicitly want the sample-ids to be relatable to the trip-id, and we also want trip-ids to be linkable if a trip becomes interrupted by events like pauses or leaving and entering the highways with rural roads in between, in case a toll collection system is used as FCD source. However, only the owner of the the corresponding OBU-id should be able to exhibit a relation between two derived trip-ids. Figure 1 illustrates which logical links should be establishable. In Figure 1, entities can be derived from one another or are relatable, if and only if an arrow is present between them. The direction of derivation coincides with the direction of the arrow.

Relying on the arguments above, we consider the following requirements as reasonable for the derivation of trip- and sample-ids from OBU-ids:

Uniqueness: Trip-ids and samples-ids are system-wide (or even world wide) unique.

Hiding: Given a trip-id (resp. a sample-id) it is not efficiently (i.e. with polynomial complexity) possible to identify the corresponding onboard unit (resp. the trip-id).

Unlinkability: Given a set of trip-ids (resp. sample-ids) a poly-bounded third party cannot efficiently decide, which of them have been derived from the same OBU-id (resp. trip-id).

Optional Linking: The above unlinkability property can be circumvented if and only if some secret information is revealed, which may be escrowed and thus cannot be released without explicit permission of the data's owner.

The concatenation of two bit-strings x, y is denoted as $x \| y$. The length of a number x in bit is denoted as $|x|$. By $x \in_R X$ we mean a random choice of x from a set X. How can uniqueness of a randomly chosen value be guaranteed without communication between parties? The answer is the uniqueness of the OBU-id, that can be attached to any random number r as OBU-id$\|r$, which then trivially renders the result unique. This trick in its most general form is called *collision-free number generation*, and achieves the first of the above properties (see [17] for a more comprehensive discussion). For the other three requirements, we can use a large portion of the standard cryptographic toolbox, if we somehow manage to build cryptosystems upon concatenations of unique values (OBU-ids) and random values, as above. Fortunately, such constructions have been done, and for convenience of the reader, are briefly summarized in the following paragraphs.

In all that follows, let q be a prime such that $q \equiv 3 \pmod 4$, and let $\mathbf{G}_q$ be a (general multiplicative) group of order q, and let $\mathbf{Z}_q$ be the group of integers modulo q.

Standard vs. Fusion Exponentiation: For using unique pairs as identifiers, we rely on the work of [18, 19], which presents a generalized form of exponentiation in $\mathbf{G}_q$ that takes pairs in the basis and the exponent. The standard exponentiation in $\mathbf{G}_q$ is a mapping $\exp : \mathbf{G}_q \times \mathbf{Z}_q \to \mathbf{G}_q$, with $(g, x) \mapsto g^x$, being the x-fold multiplication of g with itself. For the basis, $\mathbf{G}_q$ is replaced by its outer product with itself, i.e. the group $\mathbf{G}_p := \mathbf{G}_q \times \mathbf{G}_q$ with component-wise multiplication. The constraint on q allows for replacing the exponents in $\mathbf{Z}_q$ by pairs from the field $\mathbf{F}_p = \mathbf{Z}_q[X]/(X^2+1)$, which is nothing else than a complex-number like extension of $\mathbf{Z}_q$. Consequently, the elements of $\mathbf{F}_p$ can be written as pairs (a, b), similarly to complex numbers with real part a and imaginary part b. Multiplication and addition work as for the complex numbers with calculations done modulo q.

In order to distinguish elements of $\mathbf{G}_p$ and $\mathbf{F}_p$ from elements of $\mathbf{G}_q$ or $\mathbf{Z}_q$, we shall use sans-serif fonts for pairs (e.g. $\mathsf{g} \in \mathbf{G}_p, \mathsf{x} \in \mathbf{F}_p$) and normal font for scalars (e.g. $g \in \mathbf{G}_q$).

Fusion-exponentiation [19] is then defined as a mapping $\xi : \mathbf{G}_p \times \mathbf{F}_p \to \mathbf{G}_p$, where

$$\xi(\mathsf{x}) = \mathsf{g}^{\mathsf{x}} := (g_1, g_2)^{(a,b)} := (g_1^a g_2^{-b}, g_1^b g_2^a). \tag{1}$$

and $\mathsf{g} = (g_1, g_2)$. The constraint $\mathsf{g} \neq (1, 1)$ then makes ξ injective. The function (1) can be shown to obey the same laws as normal exponentiation: $\mathsf{g}^{\mathsf{x}}\mathsf{g}^{\mathsf{y}} = \mathsf{g}^{\mathsf{x}+\mathsf{y}}, (\mathsf{g}^{\mathsf{x}})^{\mathsf{y}} = \mathsf{g}^{\mathsf{x}\mathsf{y}}, (\mathsf{g}\mathsf{h})^{\mathsf{x}} = \mathsf{g}^{\mathsf{x}}\mathsf{h}^{\mathsf{x}}$. This justifies the notation g^{x} and the naming fusion-*exponentiation*.

Intractable Problems: The security of the solutions presented in this paper rests on the intractability of the following problems, which can identically be formulated using both the standard exponentiation (classical version) and the fusion exponentiation. Using the latter, the resulting problems are either computationally equivalent or maybe even stronger than their classical counterparts [19].

For formulation of the problems, assume $x, y \in \mathbf{Z}_q$ and $g \in \mathbf{G}_q$ with $g \neq 1$. There is no known polynomial time algorithm to solve any of the following problems: The *Discrete Logarithm Problem* (DLP): given y, g, find $x \in \mathbb{Z}_q$, such that $y = g^x$ (in analogy to the standard logarithm, the number x is then denoted as $\mathrm{dlog}_g(y)$). The *Diffie-Hellman Problem* (DHP) [5]: given g^x, g^y, find g^{xy}. The *Decision Diffie-Hellman Problem* (DDP) [1]: given a triple $(a, b, c) \in \mathbf{G}_q^3$ decide whether a, b, c can be written as $a = g^x, b = g^y, c = g^{xy}$ for some integers x, y. The corresponding problems in terms of fusion exponentiation are denoted as FDLP, FDHP and FDDP. The FDLP and FDHP are computationally equivalent to the DLP and DHP, respectively, but the FDDP (up to now) seems to be stronger than its classical counterpart. If for any fusion-based problem, one component of the solution (a, b) is known already (i.e. $\mathsf{y} = \mathsf{g}^{\mathsf{x}}$ and a are known parts of the solution $\mathsf{x} = (a, b)$, and only b remains unknown), then the problem is termed *half-fusion* based (denoted as HDLP, HDHP, HDDP), but surprisingly, this gives no advantage in solving the problem. The relation to the classical counterpart is the same as without additional a-priori knowledge.

Pseudonym generation: We propose using a variant of the ElGamal cryptosystem over $\mathbf{G}_p$, in which standard exponentiation is replaced by fusion-exponentiation. A user chooses the secret decryption key as $\mathsf{d} \in_R \mathbf{F}_p$, from which the public encryption key is found as $\mathsf{e} = \mathsf{g}^{\mathsf{d}}$. To encrypt an identifier $\mathsf{m} \in \mathbb{G}_p$ (containing a trip-id or OBU-id), we choose a *randomizer* $\mathsf{r} \in_R \mathbf{F}_p$ and compute the ciphertext-pair $(\mathsf{C}_1, \mathsf{C}_2) = (\mathsf{g}^{\mathsf{r}}, \mathsf{m}\mathsf{e}^{\mathsf{r}})$. To decrypt the pair $(\mathsf{C}_1, \mathsf{C}_2)$ upon knowledge of the secret key d, simply calculate $\mathsf{m} = \mathsf{C}_2\mathsf{C}_1^{-\mathsf{d}}$. This variant of the ElGamal cryptosystem is semantically secure under the assumption that solving the FDHP is hard. Moreover, given any two ciphertexts with respect to the same encryption key, it is not efficiently possible to decide if the they correspond to the same plaintext. The latter property holds due to the FDDP.

Proving ownership and relations between entities: Since a relation between any two pseudonyms (regarding a trip or sample) should only be disclosed if the owner wants it to, we can use Σ-proofs [10, 4] to avoid having to reveal secret information for that matter. In this paper we employ two standard Σ-proofs, that can *identically* be re-stated in terms of fusion exponentiation to serve our needs for using unique pairs as identifiers. Neither of these two reveals any secret information, if the verifier (FCD service provider, toll collector, etc.) behaves honestly: in [20], a Σ-proof has been proposed by which a person can prove knowledge of a discrete logarithm of a given $y \in \mathbf{G}_q$ to the base g. This can be used to prove ownership of trip-ids that are derived from OBU-ids using fusion-exponentiation.

In [3], a similar proof protocol has been given that allows for demonstrating the equivalence of discrete logarithms of two different numbers w.r.t. two bases, i.e. for given $y_1, y_2 \in \mathbf{G}_q$ and bases $g_1, g_2 \in \mathbf{G}_q$, a person can prove that $\mathrm{dlog}_{g_1}(y_1) = \mathrm{dlog}_{g_2}(y_2)$. This way, we can prove the existence of a common parent OBU-id of two trip-ids, without revealing the identifier of the onboard unit.

Using a simple trick (known as the Fiat-Shamir heuristic [7]), the above interactive proofs can be turned into non-interactive ones, allowing for a full off-line verification of assertions with no need for the owner to be available.

3. PSEUDONYM GENERATION

The problem is to let the vehicles choose their pseudonyms *independently* from each other, while maintaining uniqueness *at all times*. In the following, an approach is presented, where uniqueness of trip-ids and sample-ids is achieved by using the concept of collision-free number generation and the system-wide unique OBU-id. Furthermore, hiding and unlinkability is achieved by using one-way derivation functions based on the DLP (and related ones). Optional revocation is achieved by providing appropriate trapdoors that could be escrowed at some trusted third party.

Trip-ids: The driver chooses a secret key k at random and maintains a counter c_i (of length $|q|-1$, where q is a prime with $q \equiv 3 \pmod 4$)), which starts at any value and is incremented modulo $2^{|q|-1}$ each time a new identifier shall be generated for a trip. The i-th TID is found as

$$TID_i = \mathsf{g}^{(E(c_i,k),OBUID)}, \tag{2}$$

where $\mathsf{g} = (g_1, g_2)$. It is important to choose *two distinct* values $g_1, g_2 \neq 1$, since otherwise linking of trip-ids becomes trivial: if $g_2 = 1$, then

$$(g_1, 1)^{(E(c_i,k),OBUID)} = (g_1^{E(c_i,k)}, g_1^{OBUID}),$$

and we end up with the *same* second component for all trip-ids! Notice that the in- and output block-length of the symmetric cipher E must be $|q|-1$ bits so that c_i can be encrypted correctly and a reduction modulo q is avoided. The left part of the exponent in (2) locally randomizes the counter using the key k, while the right component guarantees system-wide uniqueness of the exponent. Injectivity of (1) (provided if $(g_1, g_2) \neq (1,1)$) ensures that the resulting trip-id inherits uniqueness from the exponent.

Sample-ids: Several sample-ids can be derived from a trip-id using ElGamal encryption with changing random inputs as follows:

$$SID_j := \mathsf{g}^{(E(t_j,r_j),r_j)} \| TID_i \mathsf{e}^{(E(t_j,r_j),r_j)}, \tag{3}$$

where t_j is the current time and r_j is a random "key", which is sufficiently long to prevent known attacks. As the time is known, the security of this ElGamal-variant holds with respect to the FDHP, as only t_j may be known, but r_j and therefore $E(t_j, r_j)$ both remain unknown. The injectivity of fusion-exponentiation and the local uniqueness of t_j ensures local uniqueness of SID_j, since $(E(t_j, r_j), r_j)$ is a collision-free number generator according to [17, 18]. On the other hand, two sample-ids derived from different trip-ids, possibly at the same time t_j with the same randomizer r_j, will always differ, since even if the exponent $(E(t_j, r_j), r_j)$ is the same for two users, the different trip-ids will make the two sample-ids different. If the time or randomizer is different, then the exponents and hence the first half of the sample-id is unique already.

4. SECURITY & PRIVACY

Hiding (Trip-ids): Extracting an OBU-id from a given a trip-id is difficult, because the left exponent of the trip-id is the output of a symmetric encryption function and hence appears random. Even if the state of the counter c_j is known, the correct key k remains unknown. If a standard block-cipher with at least 80 bit keys (to prevent brute-force attacks) is used, then this is infeasible. Also, notice that not even a trusted third party can disclose a trip-id because no easy-to-apply trapdoor can be given. Thus, a high degree of anonymity is achieved.

Hiding (Sample-ids): Extracting a trip-id from a sample-id is equal to decrypt an ElGamal-ciphertext without the secret key, and hence in turn equal to solving the FDHP or HDHP in our setting. Fortunately, the randomizer for the sample-ids is of the form $(E(t_j, r_j), r_j)$, where r_j is chosen at random, thus making the HDHP hard. The trapdoor to the generation of a sample-id is the ElGamal decryption key d that can be escrowed at a trusted third party, and can only be disclosed if the owner of the sample grants the permission to do so.

Unlinkability (Trip-ids): Given any two trip-ids, we show that it is computationally infeasible for a poly-bounded adversary to decide if both have been derived from the same OBU-id: Let

$$TID = \mathsf{g}^{(a,OBUID)}, TID' = \mathsf{g}^{(b,OBUID)}$$

be given, where $a \neq b$ with a high probability. Then,

$$\begin{aligned} TID'TID^{-1} &= \mathsf{g}^{(b,OBUID)}\mathsf{g}^{-(a,OBUID)} \\ &= \mathsf{g}^{(b,OBUID)-(a,OBUID)} \\ &= \mathsf{g}^{(b-a,0)}, \end{aligned}$$

so by definition of fusion-exponentiation,

$$TID'TID^{-1} = (g_1^{b-a}, g_2^{b-a}) \tag{4}$$

Because g_1 is a generator of $\mathbf{G}_q$, an integer w exists such that $g_2 = g_1^w$ and hence $(g_1^{b-a}, g_2, g_2^{b-a}) = (g_1^{b-a}, g_1^w, g_1^{w(b-a)})$ is a Diffie-Hellman triple. As long as w is unknown, the problem of deciding whether TID and TID' have been derived from the same OBU-id is hard under the assumption that solving the DDP is hard. If optional linking is required through a trusted third party, w could be escrowed there after the initial generation of the system parameters q, g_1 and g_2.

Unlinkability (Sample-ids): Under the assumption that the DDP is hard, the ElGamal cryptosystem provides the property that a poly-bounded algorithm is not able to decide if two given ciphertexts contain the same plaintext. In our case this property holds with respect to the FDDP resp. the HDDP (since at least one half of the randomizer is chosen at random). Thus, one cannot efficiently decide whether two sample-ids have been derived from the same trip-id. Again the trapdoor is d.

The reductions as described above show that our construction satisfies the first set of requirements as stated in Section 2. The possibility of linking identifier if required and permitted is subject of the next section.

5. LINKING TRIPS AND SAMPLES

In the previous section it has been shown that linking is computationally infeasible as long as two trapdoors are not available: the secret ElGamal-key d and $w = \mathrm{dlog}_{g_1}(g_2)$. In the following, we show that the OBU can prove ownership of identifiers and links among them without disclosing any private knowledge. Moreover, we show what a trusted third party can do, if she knows d or w.

Proof of Ownership and Linking: Proving ownership of a trip-id or sample-id means proving knowledge of a discrete logarithm in the fusion-setting. This knowledge is either the OBU-id itself and the values k, c_i (Eq. (2)), or the pair $(E(t_j, r_j), r_j)$ in Eq. (3). For the latter, it is useful to derive the randomness r_j directly from t_j by using either a pseudo-random function [9] or by encrypting t_j with another secret that is only known to the OBU, as both methods allow for reproducing the above pair upon a given timestamp t_j or period. Proving that two trip-ids or sample-ids descend from a common parent means proving the equality of discrete logarithms in the fusion setting. Since fusion-exponentiation behaves almost exactly as ordinary exponentiation, such proofs can be designed straightforward with respect to [20, 3] (cf. Section 2).

Optional Revocation: Assume the FCD service provider knows d and w. Given any two trip-ids $TID = \mathsf{g}^{(a,OBUID)}$ and $TID' = \mathsf{g}^{(b,OBUID')}$, the mutual correspondence can be decided as follows: compute $(A,B) = TID'TID^{-1}$ and check if $A^w = B$. This works, because by (4), $(A,B) = (g_1^{b-a}, g_2^{b-a})$ and $A^w = (g_1^{b-a})^w = g_1^{w(b-a)} = g_2^{b-a} = B$. If, by coincidence, $a = b$ occurs, then still $OBUID \neq OBUID'$, and in that case

$$\begin{aligned}(A,B) &= (g_1,g_2)^{(0,OBUID'-OBUID)}\\ &= (g_2^{OBUID-OBUID'}, g_1^{OBUID'-OBUID})\end{aligned}$$

and $A^w \neq B$, because $q \equiv 3 \pmod 4$ implies $w^2 \neq 1$. If $a \neq b$ and $OBUID \neq OBUID'$, we get

$$\begin{aligned}A &= g_1^{b-a} g_2^{OBUID'-OBUID},\\ B &= g_1^{OBUID'-OBUID} g_2^{b-a}.\end{aligned}$$

But then again $w^2 \neq 1$ implies

$$A^w = g_2^{b-a} g_2^{w(OBUID'-OBUID)} \neq B.$$

Given a sample-id, the corresponding trip-id can then be easily obtained through an ElGamal-decryption. Identifying further sample-ids with respect to the disclosed trip-id is generally possible for the FCD service provider, but requires a lot of computational effort. Two possibilities are:

1. Decryption of all sample-ids and identify which of them belong to the same trip-id.

2. For a given SID check if for every considered sample-id SID' the decryption of $SID'SID^{-1}$ yields $(1,1)$.

The first approach violates the privacy and must be prohibited through a privacy policy. The second approach requires about the same computational costs, but still protects the privacy of all vehicles, which are not involved.

6. MESSAGE OVERHEAD

Choosing $\mathbf{G}_q$ as a subgroup of an elliptic curve group [12] allows for using 180 bits for exponents (i.e. scalars) and 192 bits for each coordinate of the resulting point on the elliptic curve. According to [14], these bit-lengths are currently believed to make the DLP and related problems hard. Each point is given as a two-dimensional vector (X,Y), and for each X-value at most two possible Y-values exist, so the second coordinate can – except for the sign – be uniquely reconstructed from X. The decision whether $+Y$ or $-Y$ is the correct one requires a single additional bit (this technique is known as *point compression*), hence a point requires 193 bits for its representation. The output of fusion-exponentiation in this setting therefore requires 386bit, since for each component one point is necessary. An ElGamal-ciphertext hence requires $2 \cdot 386 = 772$ bit. So the length of a trip-id is 386 bit and the length of a sample-id is 772 bit. Other choices apart from elliptic curve groups are of course possible, but those may produce longer identifier

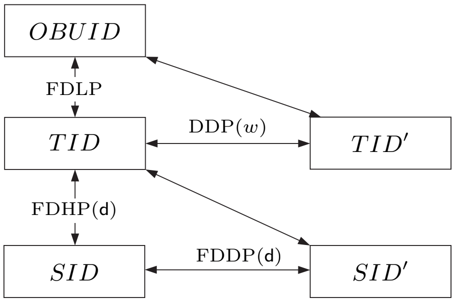

Figure 2: Computational problems (with trapdoors) protecting secret information from extraction.

7. SUMMARY AND CONCLUSION

A summary of which relations can be established is given in Figure 2, which displays the relations between entities and the computational problems whose (believed) infeasibility protects secret information. Trapdoors that allow for efficient disclosure of secret information are in possession of trusted third parties and/or the owner of an identifier, and hence privacy is protected as long as the problems remain hard, or the owner does voluntarily provide some trapdoor. In Figure 2, vertical arrows represent ownership-relations, while horizontal arrows represent the fact that two entities share a common parent identifier. Deciding on the existence of a relation is hard, unless either the problem can efficiently be solved, or a trapdoor is known. Possible trapdoors are given in brackets next to the problem. In the above diagram, these trapdoors are d, being the secret ElGamal decryption key, and w, being the discrete logarithm of g_2 to the base g_1, i.e. $g_2 = g_1^w$.

Trust is a crucial issue for a successful roll-out of an effective floating car data system. Drivers need have their privacy protected when participating in FCD projects. Under the assumption of sufficient computing power to be available in a vehicle, we have presented a method for rendering floating car data anonymous in a two-stage approach, by exploiting the identifier of a vehicles' onboard unit to create pseudonyms under which a driver can submit samples without giving the chance to be tracked by the FCD service provider. A two-stage approach for that matter creates first a trip-id from the OBU-id in such a way that a link between an OBU-id and a trip-id is always provable by the onboard unit, but the OBU-id is not efficiently recoverable from the trip-id. Based on a trip-id, samples can be related to a trip, but faked samples can be identified as such in case of doubts, by proving ownership as described above. Trip interruptions can be eliminated by relating the trip-ids to each other without revealing the underlying owner (OBU-id). Relating samples to each other upon their locations is hardly possible, since standard sampling rates for FCD vary between 10sec (cities) and 100sec (highways) [11, 8]. In cities, the number of transmitting participants should be sufficiently high in order to conceal single sources, and the vehicle density on highways should also be sufficiently large in order to allow for enough samples to hide relations between those of the same source.

Since security is a crucial issue not only for protecting value but also for acceptance by participants, related vehicle-to-vehicle or vehicle-to-infrastructure communication standards and projects (such as FLEETNET[1] for instance) may significantly benefit from our ideas too. Using an efficient and elegant generalization of standard cryptographic tools, we have presented a framework that allows for arbitrary disclosure of information within FCD systems, whilst keeping privacy of a user under his/her own control.

8. REFERENCES

[1] D. Boneh. The Decision Diffie-Hellman Problem. In J. Buhler, editor, *Proceedings of the Third International Symposium on Algorithmic Number Theory – ANTS-III*, volume 1423 of *Lecture Notes in Computer Science*, pages 48–63. Springer, 1998.

[2] Levente Buttyán, Tamás Holczer, and István Vajda. On the effectiveness of changing pseudonyms to provide location privacy in VANETs. In *ESAS 2007, LNCS 4572*, pages 129–141, Berlin, 2007. Springer.

[3] D. Chaum and T. Pedersen. Wallet Databases with Observers. In E. F. Brickell, editor, *Advances in Cryptology – CRYPTO'92*, volume 740 of *Lecture Notes in Computer Science*, pages 89–105. Springer, 1993.

[4] R. Cramer. *Modular Design of Secure yet Practical Cryptographic Protocols*. PhD thesis, University of Amsterdam, 1996.

[5] W. Diffie and M. E. Hellman. New Directions in Cryptography. *IEEE Transactions on Information Theory*, IT-22(6):644–654, 1976.

[6] Stephan Eichler. Anonymous and authenticated data provisioning for floating car data systems. In *Proceedings of the 10th IEEE International Conference on Communication Systems (ICCS)*, October 2006.

[7] A. Fiat and A. Shamir. How to Prove Yourself: Practical Solutions to Identification and Signature Problems. In A. M. Odlyzko, editor, *Advances in Cryptology – CRYPTO'86*, volume 263 of *Lecture Notes in Computer Science*, pages 186–194. Springer, 1987.

[8] F. Goessel, S. Hermann, E. Michler, and B. Wrase. Neue ansätze zur fahrzeuggenerierten verkehrsdatengewinnung. In *Kleinheubacher Tagung 2000*, 2000.

[9] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *J. ACM*, 33(4):792–807, 1986.

[10] S. Goldwasser, S. Micali, and C. Rackoff. The Knowledge Complexity of Interactive Proof Systems. *SIAM Journal on Computing*, 18(1):186–208, 1989.

[11] Frank Gössel. *Informationsentropische, spektrale und statistische Untersuchungen fahrzeuggenerierter Verkehrsdaten unter besonderer Berücksichtigung der Auswertung und Dimensionierung von FCD-Systemen*. PhD thesis, Technische Universität Dresden, Fakultät für Verkehrswissenschaften 'Friedrich List', Institut Informationstechnik für Verkehrssysteme, 2005.

[12] V. S. Miller. Use of Elliptic Curves in Cryptography. In H. C. Williams, editor, *Advances in Cryptology – CRYPTO'85*, volume 218 of *Lecture Notes in Computer Science*, pages 417–426. Springer, 1986.

[13] P. Papadimitratos, V. Gligor, and J.-P. Hubaux. Securing vehicular communications - assumptions, requirements, and principles. In *Workshop on Embedded Security in Cars (ESCAR)*, Berlin, 14-15 Nov 2006.

[14] PT. Post und Telekom Regulierungsbehörde: Bekanntmachung zur elektronischen Signatur nach dem Signaturgesetz und der Signaturverordnung. Bundesanzeiger Nr. 59, p. 4695–4696, January 2005.

[15] Maxim Raya and Jean-Pierre Hubaux. The security of vehicular ad hoc networks. In *Proceedings of the 3rd ACM workshop on Security of ad hoc and sensor networks*, pages 11–21, Alexandria, VA, USA, 2005.

[16] K. Sampigethaya, L. Huang, M. Li, R. Poovendran, K. Matsuura, and K. Sezaki. CARAVAN: Providing location privacy for VANET. In *3rd workshop on Embedded Security in Cars (ESCAR 2005)*, 2005.

[17] M. Schaffer, P. Schartner, and S. Rass. Universally Unique Identifiers: How to ensure Uniqueness while Preserving the Issuer's Privacy. In S. Alissi and H. R. Arabnia, editors, *Proceedings of the 2007 International Conference on Security & Management – SAM'07*, pages 198–204. CSREA Press, 2007.

[18] Martin Schaffer. *Collision-Free Number Generation: Efficient Constructions, Privacy Issues, and Cryptographic Aspects*. PhD thesis, Alpen-Adria-Universität Klagenfurt, Fakultät für Technische Wissenschaften, 2007.

[19] Martin Schaffer and Stefan Rass. Secure collision-free distributed key generation for discrete-logarithm-based threshold cryptosystems. In G. Dorfer, G. Eigenthaler, H. Kautschitsch, W. More, and W. B. Müller, editors, *Proceedings of the Klagenfurt Workshop 2007 on General Algebra*, Klagenfurt, 2008. Verlag Heyn GmbH & Co KG.

[20] C. P. Schnorr. Efficient Identification and Signatures for Smart Cards. In G. Brassard, editor, *Advances in Cryptology – CRYPTO'89*, volume 435 of *Lecture Notes in Computer Science*, pages 239–252. Springer, 1989.

[21] A. Stampoulis and Z. Chai. Survey of security in vehicular networks, project CPSC 534. `http://zoo.cs.yale.edu/~ams257/projects/wireless-survey.pdf`, 2007. last access: Jan 29th, 2008.

[22] T. Zhou, R.R. Choudhury, P. Ning, and K. Chakrabarty. Privacy-preserving detection of sybil attacks in vehicular ad hoc networks. In *Mobiquitous*, Aug 2007.

[1] `http://www.et2.tu-harburg.de/fleetnet/` (last access: April 9th, 2008)