PriPAYD: Privacy-Friendly Pay-As-You-Drive Insurance

Carmela Troncoso, George Danezis, Eleni Kosta, Josep Balasch, and Bart Preneel, *Member*, *IEEE*

Abstract—Pay-As-You-Drive insurance schemes are establishing themselves as the future of car insurance. However, their current implementations, in which fine-grained location data are sent to insurers, entail a serious privacy risk. We present PriPAYD, a system where the premium calculations are performed locally in the vehicle, and only aggregated data are sent to the insurance company, without leaking location information. Our design is based on well-understood security techniques that ensure its correct functioning. We discuss the viability of PriPAYD in terms of cost, security, and ease of certification. We demonstrate that PriPAYD is possible through a proof-of-concept implementation that shows how privacy can be obtained at a very reasonable extra cost.

Index Terms—Communication system security, legal factors, privacy.

1 Introduction

Insurance represents a large fraction of the cost of owning a car. In order to lower costs for both owners and insurers, insurance companies have developed Pay-As-You-Drive, PAYD, (or Pay-Per-Mile) models. In contrast to the current pay-by-the-year policy, customers are charged depending on where and when they drive, instead of a fixed premium per year. For each kilometer that a car is driven, the statistical risk of accident is calculated and translated into a personalized insurance fee. A PAYD contract clearly lays out the exact fares for driving under different conditions depending on the type of road, time of day, etc.

Pay-As-You-Drive insurance models are hailed as the future of car insurance due to their advantages for users and companies [1], [2], [3]. Arguments in favor of PAYD insurance are first, that the insurance fees applied to each user seem fairer than the ones in the pay-by-the-year scheme, as customers are only charged for the actual kilometers they travel. Customers could also reduce their monthly bill by choosing cheap itineraries or by just not using their car. This, in turn, would make vehicle insurance affordable for lower income car users (e.g., young people) or for people who wish to have a second vehicle. Second, PAYD policies can be socially beneficial by encouraging responsible driving, for instance, discouraging youngsters from driving at night. This would decrease the risk of accidents, which in turn saves money for users and insurers (aside from saving lives). Finally, PAYD has a potential environmental benefit, as it discourages driving, hence reduces energy consumption and pollution emissions. Due to all these advantages, PAYD insurance policies are supported by motorist associations like the National Motorist Association [4] and the American Automobile Association [5]; and they are being widely developed by insurance companies all over the world like Uniqa Group [6] (Austria), Hollard Insurance [7] (South Africa), MAPFRE [8] (Spain), or Aioi [9] (Japan), among others.

Although PAYD insurance seems to have many advantages, its current implementations involve an inherent threat to user's privacy. This has been one of the factors slowing adoption and has been reported as one of the reasons some major schemes were discontinued (as the Norwich Union's PAYD program [10]). In most of the implemented schemes, the full information used for billing (the time and position where the car was) is gathered by a black box in the car. It is then transferred to the insurance company and, in some of the cases, to a third company providing the location and/or data transportation infrastructure. Some companies claim to provide privacy preserving PAYD schemes, as they collect only statistics about the location data (e.g., how much time a driver was driving in a highway, but not when and on which highway). However, these statistics are computed by a third party who collects and keeps the raw location data, hence the threat to privacy does not disappear but is only shifted. As a result, insurance companies and/or third parties build vast databases of location data. For instance, Octo Telematics [11] reports to work for more than 30 companies in Europe and to have more than 1,000,000 clients in May 2010. Even if the location traces are anonymized, it has been demonstrated that the identity of the driver can be recovered from the traces themselves [12], [13] thus the privacy of the customers is still put in danger.

This situation has a downside both for the companies and the customers. For the former, managing these huge databases implies the risk of information leakage [14], [15] and consequent damage for the company in terms of cost and/or reputation toward the public. For the client the main disadvantage is that, in this model, the insurance company has the ability to track any of its users with ease and

- *C. Troncoso, J. Balasch, and B. Preneel are with IBBT-Katholieke Universitiet Leuven, K.U. Leuven ESAT/COSIC, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium. E-mail: {carmela.troncoso, josep.balasch, bart.preneel}@esat.kuleuven.be.*
- *G. Danezis is with Microsoft Research Cambridge, Roger Needham Building, 7 JJ Thomson Avenue, Cambridge CB3 0FB, United Kingdom. E-mail: gdane@microsoft.com.*
- *Eleni Kosta is with the Katholieke Universitiet Leuven, ICRI-K.U.Leuven-IBBT, Sint-Michielsstraat 6, 3000 Leuven. E-mail: Eleni.Kosta@law.kuleuven.be.*

Manuscript received 4 Nov. 2009; revised 28 May 2010; accepted 17 Oct. 2010; published online 23 Nov. 2010.
For information on obtaining reprints of this article, please send e-mail to: tdsc@computer.org, and reference IEEECS Log Number TDSC-2009-11-0162. Digital Object Identifier no. 10.1109/TDSC.2010.71.

TABLE 1
Current PAYD Implementations

Company	Country	Method to gather data	Method to transmit data	Known Patent	Third Party	Privacy invasive
Corona Direct [20]	Belgium	Odometer read yearly	Read by the mechanic	-	No	No
WGV [24]	Germany	GPS	User gives the info	-	No	No
Aioi [9]	Japan	Device in car	Radio or GSM	WO2005/08365 [26]	Yes	Medium
Progressive Casualty TRIPSENSE [27]	US	Device in car and software	User send info through internet	-	No	Medium
Toyota [28]	Japan	Device in car	Radio or wired communications	JP002259708 [29]	Garage	Medium
MAPFRE [8]	Spain	Full GPS data	GSM network	-	Yes	Medium/High
Hollard Insurance [7]	South Africa	GPS	GSM network	-	Skytrax [30] (Mobile Data)	Yes
iPAID^{TM} [31]	Canada	Device in car	USB key, Bluetooth, GPS	-	Themselves	Yes
Norwich Union [33]	UK	Full GPS data	GSM network	EP0700009 [32]	Yes	Yes
Progressive Auto Insurance [35]	US	Full GPS data	GSM network	US5797134 [34]	No	Yes
Sara [36]	Italy	GPS	GSM network	-	Movitrack [37]	Yes
STOK [38]	Netherlands	GPS	GSM network	-	Themselves	Yes
Skymeter [39]	US	GPS	Wireless	-	Themselves	Yes
Octo Telematics [11]	Europe	GPS	GPRS	-	Themselves	Yes
Coverbox [40]	UK	GPS	-	-	Themselves	Yes
Uniqua [6]	Austria	Full GPS data	GSM network	-	No	Yes

precision and even make profit out of these data [16]. The possession of an individual's fine-grained location data allow for inferences with private data, as the places visited may reveal sensitive information. For instance, affiliation to a given political party when its headquarters are visited often, or health status when the person has frequent appointments at a specific clinic (e.g., doctors specialized in AIDS treatments). Iqbal and Lim show how GPS data can be automatically analyzed to produce profiles of driver's behavior, social activities, and work activities [17]. For instance, in their study they could identify the home location of the population subject to the experiment in four out of five cases. For the last case, the error on the prediction stemmed only from the fact that the car was parked in an underground parking instead of in front or near the actual building. An extensive discussion on the consequences of losing location privacy can be found in [18].

Our contribution is to propose PriPAYD, a privacy-friendly scheme, where the premium computation is done in the car's black box, and only the minimum information necessary to bill the client is received by the insurance company. We provide an overview of our architecture, in which well-understood techniques are combined to give assurance to the user that the insurance company does not get more information than necessary, while granting him (or a judge in case of dispute) access to all the data. Our techniques also permit easy policy management and policy enforcement by the insurer. We also describe how user's misuse of the system can be detected such that appropriate measures can be taken by the insurance company.

The rest of the paper is organized as follows: Section 2 presents a survey on the current implementations of PAYD policies, and related work. In Section 3, we give a detailed description of our privacy-friendly scheme. We describe our proof-of-concept implementation in Section 4. We discuss the feasibility of our scheme and compare it with the previous work in Section 5, and finally we conclude in Section 6.

2 A Survey of PAYD Implementations

Pay-As-You-Drive plans are offered by many insurance companies around the world, gathering the data in a variety of ways. We can distinguish three types of policies, based on how privacy-invasive they are. Some of them do not imply any breach of privacy since the data about the amount of kilometers traveled (no location information) needed to compute the premium, is provided only once a year from a fixed location. The second type, despite not recording location information, collects data in geographically distributed points, which allow the insurance to estimate the movements of the user. Finally, the last model collects GPS data to track all the car's movements. In the rest of the section, we present real-world systems that fit in these three categories. Table 1 offers a summary.

The first type of systems, the least privacy-invasive, are also the least numerous. Examples of this model are Corona Direct [19] (Belgium) and Polis Direct [20] (Netherlands). They only use the data from the car odometer, obtained in annual vehicle inspections, and per-kilometer premiums

are calculated by dividing current premiums by the current policy maximum annual kilometers. GMAC Insurance [21] (US) follows a similar idea, and offers discounts for their clients based on their monthly mileage. This is hardly privacy-sensitive, since it does not reveal where the car has been over time. Although this scheme is the most advisable [22], it seems not to be economically practical. The management costs of reading the car odometer are high, thus the reduction in the fee is small and the policy benefits are negligible for both customer and insurer. For this reason, Polis Direct removed this policy from the market [23].

WGV [24], a German insurance company, offers a different scheme that does not infringe the user's privacy either. They collect the car speed and they use GPS to locate the road where the car is driving, but with the sole purpose of checking that speed limits are being observed and without saving any location data. When the speed limit for a given route is exceeded, the user collects "negative" points that have repercussions on his final premium.

Other nonprivacy-invasive methods are based on prepaid insurance, as proposed by Milimeter [25] in Dallas (US). In their scheme, "auto insurance by the mile," customers buy miles in advance and renew as needed, thus there is no need for vehicle tracking devices. The system is not yet available, and its legality under European law is questionable (see Section 5.1).

In the second group of PAYD policies, we find models such as the one from Aioi [9], a Japanese insurance company. They install a device in the car that records the odometer value, the car condition, and the time. This information is collected by receivers placed by the road, thus allowing to approximate the car trips. This datum is sent to the insurance central database for billing purposes and, also, to the database of the company that provides the data collectors.

Two companies, AVIVA (Canada) and Progressive Casualty Insurance (US), supply devices (Autograph [41] in the first case, and TripSense [27] and MyRate [42] in the second) that can be easily connected by the user to the On Board Diagnostics II (OBDII) port of the car. This device collects: trip start and end time, miles driven, duration of trip, number of sudden starts and stops, and time and date of each connection/disconnection to the OBDII port. These data can be seen by the client in a personal computer and can be exchanged for discounts if sent to the insurer. In this case, however, Progressive will retain information collected or derived from the device indefinitely. In Germany, a similar device is used by Swiss Re [43], and a variant is adopted by DBV Winterthur [44] giving the user the opportunity of exchanging data for discounts. They collect, through the use of GPS, the route information of the vehicle, from which they infer the kilometers traveled, the speed and the behavior of the user.

Some patented schemes propose models that also fit in the second group. Patent ES2108613 [32], for instance, suggests a model where the car is fitted with speed sensors and accelerometers, and also collects data from special devices on the roadside. The gathered data are sent to the insurer via "data collectors" present in garages and petrol stations.

Finally, we can find models that base their premium calculations on continuous collection of data, which leads to the gravest invasion of customers' privacy. Many insurance companies have chosen to follow this model. For example: Hollard Insurance [7] performs a PAYD insurance based on Skytrax GPS service (supplied by Mobile Data [30]) in South Africa. This GPS module is installed in the car, records all the data (position, time, speed, etc.) and stores it in a server, where the client can access it from the Internet. This is privacywise the worse model, as not only the insurance company builds a huge database of clients' location data, but also a third party has a copy of this database.

Progressive Insurance Corp. (US) [35], registered the US Patent US5797134 [34], in which they propose to gather the necessary data for billing (where, when and how much the car has been driven) using GPS. At the end of each month, a GSM phone fitted in the car (which is part of the policy) reports to Progressive all driving patterns. They go even further, proposing the collection of data that would give an idea of the safe operation of the vehicle by the driver such as speed, safety equipment used (seat belt, turn signals, etc.) rate of acceleration, rate of braking, or observation of traffic signs.

This scheme was closely followed by Norwich Union [33] in the United Kingdom, owner of European patent (EP) number 0700009 [32]. Their policy was based on less data as they only considered the time of the day, the type of road (more or less dangerous) used, and the number of kilometers driven. Nevertheless, Norwich Union kept all the location and timing data collected from the GPS signal, that was transferred to their central database via GSM. In their scheme, the collected data were handled by at least four companies: Norwich Union itself, a market agency and another two companies handling the back-end systems. After two years in the market, the company stopped the program due to the small amount of clients that signed for this scheme. One of the main reasons for this failure is that there was significant uncertainty about the protection and use of information obtained by the insurance company [10].

We find a very similar scheme in Austria, where Uniqa Group [6] offers an insurance that uses a GPS device in the car to collect location datum and transmits it once a day, via GSM, to the base station of the company. The data are then used to calculate the monthly premium of the client. The same scheme is announced in Germany by Pincar [45]. In Italy, "Sarafree Km" is offered by the insurance company SARA [36]. In this scheme, customers install a GPS device (supplied by Movitrack [37]) that collects the datum and sends it to the insurer. The company calculates from these data the client premium based on the actual kilometers driven.

Among the insurance companies providing PAYD based on satellite data, a Spanish company makes some effort to protect the privacy of their users. MAPFRE [8] offers the installation of a black box in the car that records: kilometers driven, type of roads used, average length of the trips, time of the day, regions in Spain where the car has been driven, average speed, and percentage of night hours. In order to obtain these data, the company relies on a third party to receive and process the raw data, providing the insurance only with aggregated anonymized data. In this procedure, the third party does not have access to the personal data

corresponding to the GPS data it is processing. However, full anonymity is not possible in this scheme, as the datum itself may reveal the actual identity (and other personal data) of the vehicle owner [12], [17] via simple profiling.

There are also third parties that offer insurers the necessary technology in order to implement GPS PAYD policies like STOK [38] (Netherlands). This company offers a system to be installed in cars, as well as the means to transfer the information collected to the insurance company and present it to the client (while having the data themselves). The same service is offered by iMetrik [46] in Germany.

Although Norwich Union has stopped PAYD insurance, Coverbox [40] (Wunelli Limited) has taken over in offering this kind of services. This company acts as a proxy between the user and main insurance companies in the United Kingdom. Through a black box installed in the vehicle, Coverbox monitors customers usage in terms of distance covered and the time of day or night a vehicle is in use, and computes the premium according to the kilometers driven in off-peak, peak, or "super-peak" periods. In their description of the service, it is unclear the amount of data shared between Coverbox and the insurance company that finally charges the customer.

In Europe, Octo Telematics [11] offer themselves to collect and process location data for PAYD insurance. In their scheme, the location data of drivers are collected in their central database. Then, these data are processed and aggregated information is given to the insurance company for the final billing. Further, these data can be available also for car makers and authorities. We note that, although final entities receive only aggregated or "anonymized" information, the central database holds a precise log of the system users' movements.

A more developed technology is the one introduced by iPAID [31] (Canada); they present a GPS tracking solution for driving data collection. It records when, where, how far, how fast, and how aggressively a vehicle was driven on the in-vehicle iPAID unit. These data can be transmitted to the central server in a passive way (via a USB key, Bluetooth, or wirelessly) or an active way (using the GSM network) which compiles it in statistics and trip logs, which the user can look up through the web. These statistics are also given to the insurance company in order to calculate the premium.

A comprehensive list and description of current Pay-As-You-drive practices can be found in [47].

2.1 Other Related Work

Besides Pay-As-You-Drive insurance, similar schemes are being developed in several countries for road pricing, and an European Electronic Road Tolling system is being introduced in the European Union [48]. Road pricing schemes charge motorists directly for driving on a particular road or in a particular area, and can include fuel taxes, license fees, parking taxes, tolls, and/or congestion charges. Some of these schemes are semiprivacy-invasive, like the proposal form the Maltese government in which dedicated cameras monitor and photograph the number plates of cars entering and exiting in Valleta [49]. Given the collected data, it is possible to learn when cars enter or leave the city, but not to infer all their movements. However, other governments (e.g., United Kingdom [50]) propose satellite-based schemes in which, as in PAYD, the full record of cars' movements is logged. See [51], for a comprehensive study of road pricing schemes. Furthermore, companies that are already collecting data for road tolling try to market these data for PAYD insurance use, as for instance, Skymeter [39] (US) does. They provide distance, time, place, velocity, and acceleration information depending on the selected policy requirements and affirm these can reliably be used to calculate insurance premiums.

Car telematics is not the only area where new technologies that allow fine-grained measurements increase the danger of customer's privacy violations. In the field of electrical metering, a new billing system in which the electrical consumption of subscribers is continuously monitored, is being pushed in the European Union [52]. The goal of the system is to help the service provider to optimize the electrical grid (thus improving energy use). As in PAYD, these measurements would result in a decrease on the final premium, but mining these data allows to deduce users habits (e.g., when the inhabitants are not at home). A solution was presented by LeMay et al. in [53], [54] that preserves user privacy against both the electric company and casual or malicious eavesdroppers. As in PriPAYD, the only datum transferred (and thus available to third parties) is aggregated data that can be used for billing, but does not leak any sensitive information.

2.2 The Abstract "Continuous Model"

We chose to model one of the most privacy-invasive, data-hungry PAYD model that is available today. We call "Continuous Model" any system in which the data are collected by GPS, using a black box installed in the car, and then are sent to the insurance company (directly or through an intermediary). This model is a generalization of all the other models, meaning that less privacy-invasive policies (such as those that only take into account odometer readings) can also be implemented using it.

The GPS-based Pay-As-You-Drive insurance is illustrated in Fig. 1a. It works as follows: as the car is being driven, GPS data are collected by the insurance *black box*. The full data gathered are sent to the insurance company, who will do the accounting to obtain the client's premium and send the bill by traditional post, together with a user-friendly (reduced) version of the full GPS data. (This is very close to the Coverbox [40] operating procedures.)

It is important to note that the correctness of the billing depends on the black box. For this reason, both the customer as well as the insurer have stakes in its correct functioning, as well as incentives to game it to their advantage. To prevent malicious behavior in practice, the boxes are provided by the insurance company and should be protected using tamper evidence and tamper resistance techniques [55] making it hard for the car user to modify their behavior. Moreover, the car user receives a detailed bill that allows him to audit the trips contributing to his premium and legally challenge the premium if they do not correspond with actual car movements.

In a typical model such as the Coverbox policy [40], GPS data points (coordinates and time) need to be mapped to different road categories (more or less dangerous) in order to extract the final premium. The fact that these

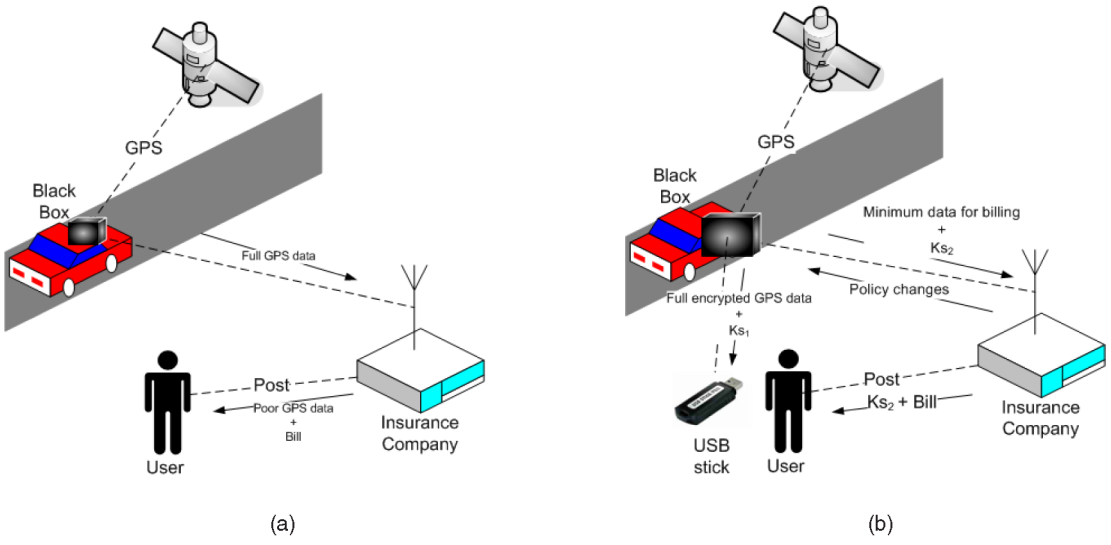


Fig. 1. (a) "Continuous Model" PAYD scheme, and (b) Privacy-friendly PAYD model.

computations can be offloaded onto powerful servers having access to online up-to-date data sources makes the "Continuous Model" very flexible and easily adaptable, as premium rate adjustments and/or road map updates can be performed centrally. We note however that the frequency of dynamic changes in the premium rates is restricted by the fact that the insurance policy should be predictable and easy to understand by the customer (which is required for a contract to be fair).

This model puts service providers (either insurers or third parties) in an advantageous business position. With all the data collected, new services (traffic information, pollution information, etc.) can be offered to customers. It also allows providers to perform data mining to detect potential fraud.

However, the obvious disadvantage of the "Continuous Model" is that it is privacy-invasive, as the data collected by the insurance company is sufficient to track almost every movement of a car over time. The data are transmitted sometimes over third parties, such as the GSM network or a third-party location data provider. Once the location data have been transmitted, the data subject has little control over it. These data could be stored or retained for long periods, as well as used for other purposes than the ones they have been collected. Although Data Protection legislation may impose limits on what can be done with it, the penalties for breaching them are often very light. So far, GSM lawful interception interfaces could be used by the authorities to get access to the location data without the knowledge of the users or even insurers [56]. Given the legal void, we fully expect to see such attempts. Moreover, after 20 years, the security of the original A5/1 and A5/2 GSM encryption algorithms has been degraded to an extent that production cryptanalysis on massive GSM traffic is within reach of many organizations [57].

3 PriPAYD: Privacy-Friendly PAYD Insurance

We present the PriPAYD architecture (see Fig. 1b) that follows closely the "Continuous Model" with the exception that the raw and detailed GPS data are never provided to third parties. The main advantage of PriPAYD, is that the insurance company receives only the billing data instead of the exact vehicle locations (thus cannot invade the user's privacy) while being sure he is receiving the correct data. The client can check that only the allowed data are being transferred to in the insurance company database and the raw data are available for the client to check the correctness of the bill in case of dispute between user and insurer.

Before diving into the details of the scheme, it is important to delineate our threat model. There is little point for our system to try to protect user's privacy beyond what road users already expect today. We assume that any third-party adversary that has extensive physical control of the car will be able to track it (by simply installing their own tracking system).

The objective of PriPAYD is to limit casual and/or deliberate surveillance by the insurance company or any third parties (with limited physical access to the car), as well as preventing the aggregation of a mass of location information in centralized databases. Fine-grained location/timing information should be hard to obtain for any third party except the policy holder, who has the right to audit the bill and ensure its fairness. This protection still allows for surveillance of the drivers (in case, they differ from the policy holders), but we are satisfied that no systemic surveillance risk is introduced beyond what is already possible today.

Our design safeguards the *privacy of the policy holder* and the *integrity of the billing information*. Yet some attacks against the availability of the PriPAYD (or previous PAYD schemes) cannot be prevented while using cheap, off-the-shelf, technology such as GPS and GSM. Our design attempts to detect that such attacks are taking place, but how they are dealt with has to be the subject of agreement between the insurance and the policy holder, and appropriate actions or penalties must be codified in the contract to deal with them. Our guiding design philosophy throughout is that the privacy-friendly mechanisms should introduce no additional vulnerabilities in PAYD with respect to the "Continuous Model."

3.1 Overview of PriPAYD

The key difference between PriPAYD and the "Continuous Model" is that all computations transforming the GPS data into billing data are performed in the vehicle black box. The data involved in the calculation of the final premium are the number of kilometers traveled, the hour of the day, the road the user has chosen, and the rate per kilometer (hour and road type) given by the insurer (following the Octo Telematics model [11]). To perform the conversion, maps have to be available to the device, and calculations have to be performed to match the coordinates with road types. These are no more complex than the operations already supported by any off-the-shelf commercial GPS navigation system.

The rates imposed by the insurer or other policy parameters can be initialized in the black box when installing it, and can be updated later in a trustworthy manner through signed updates. For the purposes of this work, we consider that policies have a unique ID_{policy} that uniquely identifies the rates interpreted by the black box. A similar mechanism can be used to perform software upgrades (uploading new firmware to the black box) with identifier ID_{code}.

Once the premium for a period of time is calculated, the amount to be paid, along with the current policy, ID_{policy}, and code version, ID_{code}, is sent in a secure way to the insurance company via GPRS, or even the cheaper SMS services (as currently done by Segurmovil [58] of MAPFRE [8]). A time stamp TS is included to protect the insurance company against reply attacks, in which a client could try to resubmit a message where the premium is low later in time. The data are signed using the black box key, and encrypted under the public key of the insurance company, in a special way that allows the policy holder to check that only the minimum billing information is transmitted (see Section 3.2).

To ensure that the black box is not acting maliciously in favor of the insurance company, we need to allow a car user or owner to audit the billing mechanism. For this purpose, we propose the use of an off-the-shelf USB memory stick. The data are recorded in an encrypted way on this token so that only the policy holder can access it, and it is signed by the black box to ensure its authenticity and integrity such that it is usable as evidence. The symmetric encryption key is generated by the black box and provided to the policy holder in two shares: one written on the USB stick and the other relayed through the insurance company and delivered by post with the bill. A simple mechanism, such as pushing a button on the box for some time, allows the encryption key to be reset. We note that certification is needed to ensure that the box properly resets this key and does not keep old information that may lead to a privacy breach in the future. See Section 5.4 for a more detailed discussion on the certification process.

3.2 The Security of PriPAYD

At the heart of the PriPAYD security policy, we have a two-level Bell-La Padula policy [59]: the confidential (high) level contains the sensors and records of the vehicle position and at the restricted (low) level we have the billing information. The only party that is authorized to access the confidential information is the policy holder, while the insurance company is only authorized to access the billing information. (Note that there is no restriction in the insurance company sending information up to confidential, i.e., policy or software updates.) In this context, transferring billing information to the insurance company is an act of *declassification*, since the data at high level are sanitized (only the amount of the final premium is sent) to not leak any information, and sent to low. The provision of the detailed location records by the policy holder, as part of a dispute, is an even more radical act of declassification.

Three key security properties are required from the channel that transfers the billing data from the vehicle to the insurance company:

Authenticity. Only the black box can produce billing data that are accepted as genuine by the insurer or any other third party.

Confidentiality. Only the insurer and the car owner should be able to read the billing data transmitted.

Privacy. The policy holder should be able to verify that *only* the billing data are sent to the insurer.

Authenticity and confidentiality. A public key signature scheme [60] can be used to certify that the data have been generated and sent by the black box. As in the "Continuous Model," the signature key in the black box is difficult to extract due to a custom tamper-resistant solution [61] or established smart card [62] technology. Public key encryption [60] can be used to encrypt the billing information (Data) under the public key of the insurer. There is no key distribution problem since the fingerprints of all public keys are seeded in the box when the device is fitted.

We denote a message sent by the black box to the insurance company,

$$\mathrm{M} = \mathrm{Enc}_{\text{Insurer Key}}(D, \mathrm{Sig}_{\text{Box Key}}(D)). \quad (1)$$

In (1), $D = (\mathrm{Data}, ID_{\text{policy}}, ID_{\text{code}}, TS = \textit{time stamp})$, where ID_{policy} and ID_{code} indicate the policy and the firmware used in the computation of Data, as explained in Section 3.1. We note that the Privacy property, that allows the user to verify that only billing data are transferred, can also be enforced. Any signature scheme ($\mathrm{Sig}_{\text{Box Key}}(\cdot)$) as well as public key encryption scheme ($\mathrm{Enc}_{\text{Insurer Key}}(\cdot)$) are verifiable: the policy holder can be convinced that the encryption is correct by being given the randomness used to perform the encryption operation (in the detailed audit log). The signature can then be verified to ensure it is correctly computed on D. Verifying these only requires the public key of the insurance and the verification key of the black box, that are public.

Privacy. The task of verifying that no other information is contained in the messages is made difficult by the existence of *subliminal channels* [63], [64] (or covert channels) in signature schemes with the potential to leak information from a maliciously programmed black box back to the insurance company. Subliminal channels, as well as techniques to limit their capacity, have been extensively studied in the multilevel secure systems literature. PriPAYD implementations should either use signature and encryption schemes that are free from such channels, or estimate their capacity and keep it under a certain threshold [65]. For instance, the client should have control over the source that

produced the randomness used in the encryption such that no message can be embedded on it (see Section 5.4). A further security measure would be to let the user choose when and where does the black box communicate with the insurance company. This measure avoids covert messages hidden in the time or location where the message was sent and has a positive influence in the privacy-preserving properties of the system (see Section 5.3). Other ways to give the user full control over the data transmitted would be to use signcryption [66] or a deterministic authenticated encryption scheme [67].

Privacy-friendly auditing. A detailed log of all the vehicle's movements (consisting of location and time) and other audit information can be extracted from the black box (signed to ensure its authenticity and that the client cannot tamper with the data), by plugging a portable device such as a USB stick on it. However, it should only be accessible to the policy holder. This is not a trivial requirement to fulfill since the black box and the policy holder need to share a symmetric key, unknown to any third party (including the insurance company). We solve the key exchange problem by having the black box generate the symmetric key and deriving two shares of it (using a secure secret sharing scheme [68], for instance, $K_s = K_{s_1} \oplus K_{s_2}$, where $\oplus$ denotes the exclusive or operation).

The first part of the key (K_{s_1}) is written to a USB stick when the system is initialized (e.g., by pushing a button in the box more than five seconds) and the second part (K_{s_2}) is relayed through the insurance company and received by the user as part of their billing sealed envelope. Both key parts are necessary to decrypt the detailed log of location data, and check its correctness. (Special software can be provided by any third party to reassemble the key parts, decrypt, read and present the detailed location logs on any commodity computer.) Through this mechanism, we ensure that only the policy holder can access these data, as neither the insurer nor any person with direct access to the car (e.g., garage mechanics) will have access to the whole key.

It may be the case (e.g., if the insurance and the mechanic collide) that both shares of the key are stolen, in an attempt to compromise the privacy of the policy holder. To avoid this, any time the black box is asked to output the encryption key, it creates a fresh pair of shares to be used to encrypt any further data guaranteeing forward security.

This mechanism is also useful if the policy holder is worried that his keys were otherwise compromised: he can force the reinitialization of the system. As explained before, upon re-initialization the black box records a fresh key share K_{s_1} on the USB stick, and sends the second fresh share K_{s_2} to the insurance company. To ensure forward secrecy, the old keys and past audit data are securely deleted from the box (see Section 5.3).

We note that, if the key K_s is not refreshed often enough, it may be used to encrypt a fair amount of data. This can be exploited by someone with physical access to the box, hence to the ciphertext, to decrypt and obtain the locations a vehicle has visited, or even to the guess the key K_s [69]. To avoid this problem, we propose the usage of a session key that varies over time in such a way that an attacker would never have enough data to mount this kind of attack. The frequency with which the session key needs to change depends on its length and the encryption algorithm chosen. The session key K' used to encrypt a given set of locations can be given to the user encrypted under K_s making sure that he is the only one able to recover K' while guaranteeing that a small amount of data is encrypted under the long-term key K_s.

Detection of the black box's inputs tampering. Even if the insurance company can verify the authenticity of the data and can trust the black box for correctness, once the box is installed in the car, the company has no control over its environment. A malicious client may try to take advantage of the situation and tamper with the incoming and/or outgoing signals (GPS, GSM, etc.) to reduce the final premium.

Given the difficulty of preventing attacks on technologies such as GSM or GPS, our approach consists in focusing on the detection of such attempts. In the following, we enumerate possible attack scenarios on these interfaces, and we propose technical solutions that can be implemented in PriPAYD. We note that these threats are common for any PAYD model using GPS and GSM technologies, hence the proposed countermeasures should not increase the costs of deploying PriPAYD with respect to the "Continuous Model."

The first point of attack against our system could be the manipulation of the GPS signal received by the box in such a way that the total number of kilometers or the type of roads used result in a smaller premium. Kuhn [70] describes how the user could realistically try to tamper with the GPS signal or receiver, and proposes a countermeasure. The solution offered by Kuhn relies on modifying the whole Global Positioning Signal system, therefore is unrealistic in the short term and cannot be used for our application. We propose here two solutions that, although they are not suitable for general purposes, solve the problem for PriPAYD at a reasonable cost.

The first approach assumes that the insurance company has knowledge of the car odometer value (as we argued in Section 2 for Polis Direct [20] and Corona Direct [19], this is not sensitive datum, as it is only aggregated data and does not reveal the location of the car). The total number of kilometers driven is also computed from the GPS signal information when calculating the premium and stored in the black box. This value is sent along with the billing information to the back office at the end of each billing period. Then, in regular inspections, the insurance company can check that the value computed from the GPS signal corresponds to the one captured by the odometer. In case these values considerably differ, the insurer can infer an attempt of cheating in the client side has happened and act in consequence (e.g., charging the client with a higher premium than the one he would obtain without cheating, according to the terms of the contract). However, the user shall be given the right to contest the assumption made by the insurance company. In order to discourage users from tampering with the odometer value before an inspection, its value should be also checked in case of accident and if the control fails, the insurance would hold the car owner liable for breach of his contractual obligations.

A second option to ensure GPS signal correctness would be to use the GSM localization features of the SIM card contained in the black box. In this case, the test consists in checking that the GPS coordinates received by the black box are matched by the location of the GSM cell in which the

SIM is transmitting. In order to make this possible, subtle modifications in the PriPAYD scheme are needed, as in the current description the insurance company never learns the location of the car, hence it cannot carry out the test. To allow this test, the black box would send the current GPS position to the insurer at random points in time (chosen by the box or upon request from the back-end office). Then, the company would check (together with the GSM provider) that the coordinates are actually located in the GSM cell used for the communication. In order to maintain the privacy-friendly properties of PriPAYD checks should only be allowed as very rare intervals (e.g., a few times a year), so that it is impossible to estimate the car's movements. A passive variant of this approach can also be used: the GSM subsystem can provide the black box with the signal strength of nearby mobile phone antennas. A simple database matching antennas with locations can then be used to cross-check the correctness of the GPS signal. This does not require the transmission of any position information, as the validation happens within the car. On the downside, this option relies upon the availability of a map of mobile base stations (a service that is already commercially available today).

The GPS signal correctness is not the only issue that stems from the fact that the black box is most of the time in a hostile environment. Even when the user cannot modify the GPS signal, he could try to decrease his monthly premium by blocking it part of the time, thus appearing as having driven less. For this purpose, he could break off the GPS antenna, or enclose it in a Faraday cage. He could also jam the GPS satellite signals with a close by, or even touching, transmitter. Even if those attacks are beyond the technical capabilities of most people, the technical know-how could be built into easy to download software or disseminated through rogue mechanics. Misleading GPS has other nefarious applications, such as stealing GPS tracked cargo or fooling home detention bracelets so it is likely that crooks will invest in defeating it in the near future.

A possible cheap countermeasure against GPS attacks is for the black box to measure the time the car is in movement and compare it with the GPS reading availability. Movement can be inferred by connecting the black box to the engine, which is already done in some PAYD models today. An even cheaper and more reliable way to detect movement is to include in the black boxes tamper-resistant envelope a three-dimensional accelerometer [71] to detect movement.

If GPS availability goes under a preestablished threshold, the black box can inform the company that, in turn, would take the appropriate measures (see Section 5.1). When establishing this threshold, the company should consider the fact that even when the client is not misbehaving the signal may not be received due to weather conditions, terrain features, tunnels, etc.

Another manner of tampering with the premium, would be to block the incoming and outgoing GSM signals [55]. If the former is stopped, no software nor policy updates would reach the box, therefore, no increase in the fees nor patches would be registered, resulting in an advantage situation for the user. Blocking the latter, the user would prevent his premium from arriving at the insurance company. Both cases could be attributed to failures in the GSM network or in the box itself. As with GPS signal blocking, tampering can be detected, for instance, by making sure that regular billing messages are received, using the ID_{policy} field to check that the policies are updated, etc. Once GSM jamming is detected, the user will in any case be liable for violating his contractual obligations toward the insurance company (see Section 5.1).

4 Implementation

In road pricing schemes, it is often argued that by providing the black box with the intelligence to compute the fee the system becomes more expensive and less flexible than with a back-end server. For example, in an study funded by the Dutch government [72], it is claimed that the advantages of a model based on a back-end server can only be achieved by a black box-centric solution "with very high risks and probably much higher costs." Our black box demonstrator aims at showing that the functionality of PriPAYD can be achieved within reasonable production costs, while proving the validity and correctness of the PriPAYD design.

In this section, we introduce the hardware and software modules that comprise our embedded platform and we present our results, in terms of computational load of the main operations, for a given test case scenario. A more detailed description of this prototype can be found in [73].

4.1 Modules

The PriPAYD black box presented in this paper requires the following hardware and software elements:

- A *processing unit*, which stores and runs the software executing the basic operations of the system. We choose to implement our demonstrator on the widely used ARM7TDMI [74] 32-bit low-power architecture, more precisely on an NXP LPC2388 [75] microcontroller. This microcontroller has a 32-bit RISC architecture, it can run at 72 MHz, it offers 512 kB of on-chip program memory and 98 kB of internal SRAM. For the development of applications, we make use of the Keil MCB2388 evaluation board [76].
- A*GPS receiver*and a*GSM modem*, where the former collects the location data and the latter allows to establish communications via the GSM/GPRS network. We use the Telit GM862-GPS [77] module since it combines both capabilities in a single device.
- An *external nonvolatile memory*, which stores large static parameters of the system such as the digital road map and the encrypted journey logs. The MCB2388 board offers two possible interfaces for external memory storage: an SD card interface and a USB Host port.
- A *digital road map*, a database of pairs latitude-longitude where each of these entries has assigned a type of road. In our implementation, we make use of OpenStreetMap [78] free digital road maps.

4.2 Performance

The main bottleneck of the scheme is the map-matching operation, i.e., the translation from GPS location data to type of road. This is because the GPS receiver provides (by default) a location string every second, thus the map-matching operation must be carried out within this time span.

In order to extract the timings of all operations, we run the following test case scenario:

1. We store in an SD card a digital road map database of Belgium containing around 1.5 million entries together with the insurer's policy table, which specifies a price per Km depending on two factors: type of road and time zone.
2. We drive for 40 minutes, such that different types of roads are involved. Every time that a new location string is available from the GPS receiver the map-matching operation is called in order to search which node of the digital road map database is closest to the position of the vehicle. When a new node is found, the distance between this node and the previous one is computed and stored along with the type of road and the time of the day. Note that these values together with the insurer's policy are required in order to compute the premium.
3. When the journey ends (e.g., when the user returns to his home location) the premium calculation operation is executed, returning as a result the amount of money that the user has to pay.
4. After the premium is computed the log corresponding to the journey (in our case, GPS coordinates and timing information) needs to be encrypted and stored, so that it can be later extracted by the policy holder. We implement the CCM [79] mode for authentication and confidentiality using AES [80] with a key length of 128 bits. In order to avoid the possibility that a malicious user tampers with the encrypted logs, we also generate a signature of the GPS location data with the 2,048-bit private RSA key of the box, by using the RSA-PSS [81] signature scheme.
5. Finally, we create the message to be sent to the insurer with the fee as shown in (1). The RSA-PSS signature scheme is used again together with the RSA-OAEP [81] encryption scheme. These methods make use of random padding encryption schemes to achieve a higher level of security, hence a true random number generator should be used (our proof-of-concept implementation uses a software pseudorandom number generator).

The performance of the main operations and routines when running the previous test case are presented in Table 2. The map-matching operation requires (in the worst case) around 24 million cycles to find the closest entry, which corresponds to 0.328 seconds when the microcontroller's clock is set to 72 MHz. This means that the microcontroller can run at lower clock speeds (hence reducing the power consumption) while still meeting the requirement that the map-matching of one location point is done before the next one arrives.

We note that our the map-matching operation is not optimal, i.e., we have implemented a simple algorithm based on the bisection method. The number of entries read at each search could be decreased by using smarter algorithms and more optimal digital maps, hence reducing the computation time. Our main purpose is to show that, even in our basic implementation, the microcontroller fully supports the requirements of the PriPAYD system.

In the operations carried out at the end of the journey, we see that the computation time of the premium calculation is negligible compared to the encryption and generation of the signature, which takes around 8 seconds. We note that the encryption operation depends on the length of the journey, e.g., an 80-minute journey would require around four seconds.

TABLE 2
Timings of the Implementation

Operation	Routine	Number of cycles	Time) (at 72MHz)
Map-matching (max.)		23 648 336	0.328 s
Premium calculation		4 222 008	0.058 s
Encrypt journey		579 087 068	8.042 s
	RSA-PSS	433 440 512	6.02 s
	CCM mode	145 646 556	2.022 s
Encrypt premium		462 977 056	6.430 s
	RSA-PSS	433 440 512	6.02 s
	RSA-OAEP	29 536 544	0.41 s

Finally, the last rows in Table 2 refer to the premium encryption operations. We must stress here that this operation does not need to be done at the end of each journey. We assume that there exists a contract between the policy holder and the insurer which specifies how often the total premium needs to be transmitted to the insurer, and the intermediate premium results can be then just aggregated. This operation can be assumed to be independent from the length of the journeys, i.e., the fact that the amount to pay is small or large does not affect the computational cost.

5 Discussion

5.1 Legal Considerations

PAYD insurance based on the "Continuous Model" transmits the full GPS data of the car's location to the insurance company for the calculation of the premium. As we have already explained above, it is not necessary that the server has knowledge over all these data for the accomplishment of the intended purpose, namely offering PAYD insurance. Such a "Continuous Model" was examined by the French Data Protection Authority (CNIL), which made some interesting considerations on the compatibility of such a system with the European data protection legal framework. By using the CNIL opinion as a starting point, in this section, we elaborate on the advantages of the PriPAYD from a legal point of view and we also present some potential limitations.

The CNIL dealt with the insurance policy that wished to be introduced by the insurance company MAAF Assurances S.A.[1] MAAF Assurances S.A. planned to launch a new insurance policy for young drivers, according to which the latter would agree not to drive during the weekend at night or longer than two hours, as well as not to exceed the speed limit. To check compliance with the policy, the insurance company would collect data related to the car's location, speed, type of road, hours and driving duration, and transmit them every two minutes.

1. CNIL, Délibération 2005-278 du 17 novembre 2005, portant refus de la mise en oeuvre par la MAAF Assurances SA d'un traitement automatisè de données à caractère personnel basé sur la géolocalisation des véhicules. (Opinion 2005-278 of 17 November 2005, refusing the data processing by MAAF Assurances SA based on vehicles' localisation).

The CNIL refused its authorization for the processing of the data. It argued that via the proposed system the insurance company would get information about violations of the speed limit. Such processing would involve sensitive data and would infringe Article 9 of the French Data Protection Act, according to which private entities are not authorized to process data relating to criminal offences. The CNIL further argued that the monitoring of data that reveal the movements of the car and consequently its driver with the exclusive aim the control of the respect of the contractual obligations of the driver infringes the principle of proportionality and the European principle of free movement of persons.

Following the recommendations of CNIL, the solution that has been later developed in order to launch PAYD in France involves a trusted third party, who collects the user data and sends the insurance company aggregate statistics on the journeys made by the driver. The use of a third party seemingly reduces the privacy issues discussed above, as the insurance company collects now less information about the users. However, this solution does not solve the problem of excessive collection of personal information. The collection of the data from the third party may still constitute a violation to the privacy of the driver, as sensitive information can be derived from these data if additional measures are not taken aiming at safeguarding privacy.

On the other hand, the approach taken in PriPAYD is more privacy-friendly, as it does not allow the monitoring of the car and its driver and it limits the processing of the data to the absolutely necessary for the provision of the PAYD insurance, in full respect of the principle of proportionality. Further, in PriPAYD, the service provider processes only the relevant and absolutely necessary data, in compliance with the fundamental privacy principle of data minimization. In order to bill the users, the insurance company does not need to know the detailed route a car has followed, nor the exact time of a journey. The PriPAYD model allows the insurance company to learn the clients' premium, while it avoids the collection of any other tracking data. Although additional information is stored on the black box, it remains under the sole control of the user, at the client side. In this way, the fundamental principles of proportionality and data minimization are respected.

The necessary data for billing are transmitted to the insurance company and to the other parties involved for the provision of the PAYD service. This defines the purpose for which the data are processed and any further processing must be compatible with it. In the "Continuous Model," the abundance of collected data may tempt companies to further process it, even in ways that cannot be considered compatible, to gain some business advantage. Such processing of data that can be linked, with no excessive effort, to the original user, is not allowed. Full anonymization of the data would allow any processing, as the data would not qualify as personal data any more, but such anonymization is very difficult, if not impossible [12], [17].

The insurance company is allowed to keep the data for the time period needed for the calculation of the insurance premium and during which the bill can be disputed.[2] After this period of time, the insurance company has the obligation to delete any collected data. A similar obligation exists for the mobile operator that has to delete the data after the provision of the service.[3] However, the mobile operator falls also under the scope of application of the data retention directive,[4] according to which the operator has to retain specific categories of traffic and location data for a period between six months and two years and have them available for law enforcement purposes. It should be pointed out that such an obligation does not exist for the insurance company, as the directive creates an obligation only for providers of publicly available electronic communications services or of public communications networks.

A crucial issue for the privacy offered by PriPAYD is the ownership of the recording of the black box; one of the principal goals of the model is that the company does not get data that are not necessary for charging the users. The full content of the black box with the detailed information that can serve for the audit of the bill on behalf of the policy holder must only be accessible to him. It is therefore important that the contract between the insurance company and the policy holder clarifies that, even if the actual black box belongs to the company, the contained records belong exclusively to the user. In case, this cannot be guaranteed, it is safer for the ownership of the box to be transferred to the user from the moment of its installation in the car.

As already mentioned above (see Section 3.2), the owner of the car is also expected to make proper use of the black box and to respect his obligations, mainly outlined in the contract with the insurer. He is not allowed to delete the data either from the black box or from the computer, where he has kept a copy of it, before the time elapses, during which the bill can be contested and the payment pursued. It cannot be excluded that the car owner does not allow the proper recording and transmission of the data that will affect his final charging. He could, for instance, tamper with the black box, block its programmed updates or shield the box in such a way that it does not transmit. In all aforementioned cases, his behaviour will be considered as a breach of his obligation to make fair use of the box and will be held liable. The insurance company will have the right to impose the penalties foreseen in the contract and further seek legal relief. In specific cases, the insurance company has also the right to ask the car owner to adhere to one of the regular insurance schemes or to ask the termination of the contract. In the latter case, the insurance company shall give the car owner the necessary time frame to conclude a new contract with another insurance company.

A potential limitation of PAYD may appear in cases when the policy holder and the driver of the car do not coincide, as for rental cars or company vehicles. In such cases, the driver should be informed about the presence of the black box and its functionality. Especially, in the case of rental cars, this should be explicitly mentioned in the car rental contract. When company cars are used by employees during their working hours, they should also be informed about the installation of the system. It is still an open question whether a company has the right to choose a

2. Article 6 (e) data protection directive.

3. Article 9 ePrivacy directive.

4. Directive 2006/24/EC of the European Parliament and of the Council on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC, Official Journal L105, pp. 54-63 (March 15, 2006).

PAYD insurance for a company owned car, which is used by an employee outside working hours. Current practices require the consent of both the employer and the employee.

5.2 Cost

PriPAYD does require more computations and mapping data in the black box than the "Continuous Model." Yet these are comparable to what current commercial GPS navigation systems do. Since the "Continuous Model" already relies on tamper resistance for security no additional costs should be expected from this either. Given that a processing unit with GSM and GPS interfaces is required for any PAYD model, the only additional hardware required in the PriPAYD black box is an external memory module (i.e., SD card in our demonstrator), which should not considerably increase the production costs.

Another source of costs is GSM communications and the PriPAYD model should be cheaper since less data are transmitted. Billing data can be aggregated to reduce those costs further. Updates containing new rates, maps, and policies, can be pushed to the black box either through the GSM communications or during the servicing of the car. It has to be taken into account that the extra amount of updates associated to PriPAYD with respect to current implementations, fees, and map updates (firmware updates are needed in both schemes), is likely to be small. It is reasonable to assume that the frequency with which fees are recalculated is very low (for instance, Polis Direct did this each year) and so is the rhythm with which new roads are constructed and ready for usage.

The PriPAYD design keeps the trust infrastructure to a minimum, and particularly does not require a public key infrastructure. The identity infrastructure is based on the preexisting relationship of the policy holder with the insurance company that is used as part of the key distribution mechanism. Hence, there is no cost associated with either of these.

In order to guarantee the integrity of the billing information, we have presented a series of detection mechanisms for fraud detection that could be used to trigger legal recourses with customers. Although these processes are expensive for the insurance company, they are not a consequence of introducing privacy protection in the system. Privacy-invasive PAYD implementations, as the "Continuous Model," suffer from the same vulnerabilities as PriPAYD when it comes to the integrity of the inputs and outputs of the black box. Thus, no additional cost in terms of court costs is expected in PriPAYD with respect to other implementations.

Finally, one has to take into account the cost of development and maintenance of the infrastructure. The technology and cryptography used is available off-the-shelf and developing PriPAYD should not be more expensive than the "Continuous Model." The additional engineering that is required for building a slightly more complex black box should be more than balanced by the reduced costs of the back-office systems, since they handle less, as well as less sensitive data.

Yannacopoulos et al. [82] propose a model to estimate the viability of the use of privacy enhancing technologies to protect data. Their approach is to model how much do clients value their privacy, and use this to compute the loss a company would suffer if a breach of privacy occurs and its clients need to receive compensation. This way, a company can calculate the hypothetical savings achieved through the use of privacy enhancing technologies.

5.3 Strengthening Privacy

Some additional privacy concerns should be tackled as part of a real-world implementation.

One needs to ensure that past location information can easily be deleted. We would advise implementers to never automatically store encrypted GPS data from the audit record; and users to keep this, or key material, only on the USB stick to which they were written by the black box. This allows the user to easily destroy the data by destroying or deleting the USB stick. The downside of having a token that can be destroyed as an easy and intuitive operation (a better paradigm for destroying the private data than electronic equivalents) is that, once audit records of the detailed locations have been deleted, it is difficult to challenge any bills that seem incorrect. In case, this is a user's main worry they can back up their records by simply copying the files to a computer or other storage.

It could be sustained that in order to allow the user a kind of guarantee for his privacy, he could be allowed to delete the detailed record at the cost of not being able to contest the correctness of past premiums (unless he had previously recorded a copy of these records). However, such an approach cannot be accepted unconditionally. The retailed record shall be retained up to the end of the period during which the bill may lawfully be challenged or payment pursued. Any attempt on behalf of the user to delete the data before the expiration of this period will hinder the insurance company from checking the validity and the content of the data. In case the user intentionally deletes the data, he will be held liable for violating his contractual obligations toward the insurance company. Nevertheless, in case the deletion of the data is caused due to a technical deficiency of the black box, then the liability lies with the manufacturer.

A further concern is the use of GSM to transfer the data back to the insurance company. In our scheme, the billing data does not contain any sensitive location information, but an active GSM device registered in the network does leak the cells the car is visiting. Hence, it is prudent to keep the GSM system powered down at all times except when transmitting. The transmission time and location must be chosen to minimize location leakage because of the GSM technology. Defining and using a preferred known "home" location, recorded in the box when initialized, should easily address this concern. Still, a timer in the black box should ensure that, even if the car is not present at this location for a long period of time (e.g., long trip), the monthly premium is sent to the company.

5.4 Certification and Independent Monitoring

The key objective of our design is to *not require* a trusted black box to guarantee user's privacy. This is an important requirement: the black box is commissioned by the insurance company and the user has only a limited capacity to discern its functioning. Furthermore, independent certification of even simple devices (such as the black box

described) is expensive and appropriate certification criteria have hardly been established.

Our design choice is to allow policy holders to have a full view of the output of the black box and to ensure that only the minimum billing information is transmitted. One option is to allow a device (again a USB mass storage device would be sufficient) to record all data sent between the black box unit performing the calculations, and the GSM subsystem that relays all the information back to the insurer. This solution is not invulnerable to a maliciously programmed black box that only reveals part of the conversation. On the other hand, it makes certification easier, since only a trivial property needs to hold: that all data transmitted using GSM are also recorded on the auditing device. A second approach, that offers stronger guarantees, is to physically separate (and shield) the black box from the GSM transmitter, and link them with a recording device controlled by the user. This device would record all traffic, and allow the users to verify that the data transmitted only contains the billing information. Recording cables could be sold by multiple manufacturers, or provided by privacy advocacy groups or data protection authorities.

The need to keep the security function of the black box simple to facilitate verification, has guided our choice of a one-button reinitialization mechanisms over more complex access control to the data in the box. An alternative mechanisms would be to require a PIN to be entered on the black box to access the encryption keys or audit log. This would make the operation and certification of the box more complex, and the black box more expensive.

We note that trusting the black box is still necessary for correct billing, and that the encrypted audit trail can be used to check bills or dispute them. Without third-party certification, it is impossible to ensure that the black box is not recording precise location data with the intent to provide them to a third party. Since such a box has no way of transmitting the recorded data over the air, physical access would be required to extract the data, making it difficult to turn this weakness into a mass surveillance tool. This is a known open problem [83], and physical access would require additional certification.

The certification goals for the box to provide high grades of assurance are:

- The random number generation should be based on a physical source of randomness. A pseudorandom number generator with a seed known to the insurance company would produce predictable encryption keys, leaving the audit logs unprotected. An alternative strategy would be for a device controlled by the user to be able to set the initial state of the random number generator.
- The deletion operation of the keys and the data upon reinitialization of the box should be effective. Otherwise the forward security property cannot any more be guaranteed, since an adversary may be able to get access to keys and logs from previous epochs. An alternative could be for the box to not hold any nonvolatile memory, aside a removable memory chip—that the user can physically remove and discard to preserve privacy.
- A thorough side channel analysis is necessary to ensure that the black box does not leak or transmit information through any other means than the audited GSM transmission. Enclosing the black box into a Faraday cage, using a conductive cover, could ensure this. Yet the GPS antenna, as well as the GSM module should be outside the enclosure.
- Finally, the correct implementation of the PriPAYD procedures should be certified: the black box only records the premium payment information; all raw location information is stored only in an encrypted form using the appropriate keys; and the reinitialization mechanisms works as advertised. This is only required to protect against adversaries with local access, since auditing ensure that no personal data are transmitted remotely.

As always certifying a product to such a level is a challenge, particularly since, for security reasons, the insurance company should be able to update the software to patch bugs. Two options are available to make this possible. The first is that the full update can be signed by a certification authority, after evaluation of the new features. Such reevaluation is expensive, and might slow down the deployment of security critical updates. The second option is for the software to be built in such a way that it cannot violate the key security properties as described in the previous paragraph: it does not have the interfaces to store unencrypted data or signal to the outside world in any other channel but the audited one. In this way, the original software and updates are sandboxed, and cannot violate the key properties required by PriPAYD, requiring only the infrequent certification of the sandbox.

6 Conclusions

Pay-As-You-Drive insurance policies, due to their advantages, are bound to gain popularity or even dominate the car insurance market. However, their most advanced current implementations show a fundamental disregard for the privacy of car owners, which might even slow or limit their deployment. Our survey of existing systems and practices sadly documents a move toward more, not less, privacy-invasive systems.

PriPAYD is a system that can support the deployment of very fine granularity PAYD policies while also providing strong privacy guarantees. Its core security architecture is based on simple and well-understood multilevel security components, that have been the subject of extensive study in the field of computer security since the 1970s. The PriPAYD architecture relies (as previous systems) on secure hardware for correct accounting, but privacy properties can be checked without relying on its correctness, just by auditing its output. This separates correct accounting from privacy concerns, allowing black boxes to remain fully under the control of insurance companies, while users can be sure that none of their location data are leaking. Our approach follows the paradigm of many security metering systems used for electricity or gas distribution that only record aggregate use.

There is no component or infrastructure required by PriPAYD that would make it much more expensive than current systems, as we demonstrate with our implementation. One could in fact argue that in the long run running

PriPAYD as any other privacy enhanced technology, is cheaper than privacy-invasive systems. The cost of protecting private data stores is often overlooked in the accounting of costs, as is the risk of a single security breach leaking the location data of millions of policy holders [14], [15]. In addition, PriPAYD keeps sensitive data locally in each car, in a simple to engineer and verify system. Requiring off-the-shelf back-end system to provide the same level of privacy protection to masses of data would make them, not only prohibitively expensive, but simply unimplementable.

ACKNOWLEDGMENTS

This work was supported in part by the Research Council K.U. Leuven: GOA TENSE, the IAP Programme P6/26 BCRYPT of the Belgian State, by the Flemish IBBT NextGen-ITS project, by the European Commission under grant agreement ICT-2007-216676 ECRYPT NoE phase II, and by K.U. Leuven-BOF (OT/06/40). The material in this paper was presented in part at Workshop on Privacy in the Electronic Society 2007, Alexandria, Virginia, October 2007, and at Design, Automation and Test in Europe 2010, Dresden, Germany, March 2010.

REFERENCES

[1] T. Litman, "Distance-Based Vehicle Insurance Feasibility, Costs and Benefits," technical report, Victoria Transport Policy Inst., http://www.vtpi.org/dbvi_com.pdf, 2007.
[2] F. Zahid and C. Barton, "Pay Per Mile Insurance," technical report, Davenport Univ., 2004.
[3] F. Kelly, "Road Pricing: Addressing Congestion, Pollution and the Financing of Britain's Roads," *Ingenia,* vol. 29, pp. 34-40, 2006.
[4] Nat'l Motorist Assoc. "NMA's Position on Auto Insurance," http://www.motorists.org, 1998.
[5] Am. Automobile Assoc., http://www.aaa.com/, 2009.
[6] Uniqa, http://www.uniqa.at/uniqa_at/, 2011.
[7] Hollard Insurance, http://www.payasyoudrive.co.za/, 2011.
[8] MAPFRE, http://www.ycar.es/, 2011.
[9] Aioi, http://www.ioi-sonpo.co.jp/, 2011.
[10] "Surveillance Fears Force Norwich to Scrap PAYD Car Policies," http://www.independent.co.uk/news/business/news/848562.html, June 2008.
[11] Octo Telematics S.p.A., http://www.octotelematics.com/solutions/insurance-telematics/, 2011.
[12] J. Krumm, "Inference Attacks on Location Tracks," *Proc. Fifth Int'l Conf. Pervasive Computing (Pervasive),* pp. 127-143, 2007.
[13] P. Golle and K. Partridge, "On the Anonymity of Home/Work Location Pairs," *Proc. Seventh Int'l Conf. Pervasive Computing (Pervasive),* H. Tokuda, M. Beigl, A. Friday, A.J.B. Brush, and Y. Tobe, eds., pp. 390-397, 2009.
[14] "Alert as 170,000 Blood Donor Files are Stolen," http://www.independent.ie/national-news/alert-as-170000-blood-donor-files-are-stolen-1294079.html, Feb. 2008.
[15] "Norwich Union Life Fined 1.26m," http://www.inf-sec.com/news/071217_norwich_union.html, Dec. 2007.
[16] "Big Brother is Keeping Tabs on SatNav Motorists," http://www.dailymail.co.uk/news/article-483682/Big-Brother-keeping-tabs-satnav-motorists.html, 2011.
[17] M.U. Iqbal and S. Lim, "An Automated Real-World Privacy Assessment of GPS Tracking and Profiling," *Proc. Second Workshop Social Implications of Nat'l Security: From Dataveillance to Uberveillance,* pp. 225-240, 2007.
[18] A.J. Blumberg and P. Eckersley, "On Locational Privacy, and How to Avoid Losing it Forever," technical report, Electronic Frontier Foundation, http://www.eff.org/wp/locational-privacy, 2009.
[19] Corona Direct, http://www.kilometerverzekering.be/, 2011.
[20] Polis Direct, http://www.kilometerpolis.nl/, 2011.
[21] General Motors OnStar, http://www.onstar.com/us_english/jsp/index.jsp, 2009.
[22] T. Litman, "Pay-As-You-Drive: Recommendations for Implementation," technical report, Victoria Transport Policy Inst., http://www.vtpi.org/payd_rec.pdf, 2008.
[23] Polis Direct stopt met KM Polis, http://www.autokompas. nl/nieuws/2007/05/Polis-Direct-stopt-met-KM-Polis.html, May 2007.
[24] WGV, http://www.wgv-online.de/index.htm, 2011.
[25] Milemeter Inc., http://www.milemeter.com/, 2011.
[26] Aioi, "Telematics Insurance System," 2009.
[27] Progressive Casualty Insurance, TripSensor, https://tripsense.progressive.com/, 2009.
[28] Toyota Motor Corporation, http://www.toyota.co.jp/, 2011.
[29] S. Nakagawa, K. Mori, A. Shinada, K. Nunokawa, H. Okajima, and M. Sasaki, "Vehicle Insurance Premium Calculation System, On-Board Apparatus, and Server Apparatus," Mar. 2001.
[30] Skytrax, https://www.skytrax.co.za/index.asp, 2009.
[31] iPAID, http://www.ipaid-insurance.com/, 2009.
[32] S.M. Perez, "Individual Evaluation System for Motorcar Risk," Dec. 1997.
[33] Norwich Union, http://www.norwichunion.com/pay-as-you-drive/, 2011.
[34] M.R. John, C.A. Dean, and H.J. Patrick, "Motor Vehicle Monitoring System for Determining a Cost of Insurance," Aug. 1998.
[35] Progressive, http://www.progressive.com/, 2011.
[36] Sara Assicurazioni, http://www.saraassicurazioni.it/, 2011.
[37] Movitrack, http://www.movitrack.it/, 2011.
[38] STOK, http://www.stok-nederland.nl/, 2011.
[39] Skymeter Corp., http://www.skymetercorp.com/, 2011.
[40] Coverbox Wunelli Limited., http://www.coverbox.co.uk/, 2011.
[41] Autograph, https://secure.avivacanada.com/autograph/product.php, 2009.
[42] Progressive Casualty Insurance, MyRate, http://auto.progressive.com/progressive-car-insurance/myrate-default.as, 2011.
[43] Swiss Re, http://www.swissre.com/, 2011.
[44] DBV Winterthur, http://entry.dbv-winterthur.de/, 2009.
[45] PINCAR AG, http://www.pincar.de/, 2011.
[46] iMetrik, http://www.imetrik.com/, 2011.
[47] S. Minguijon-Perez, "Pay As You Drive Directory," http://terra.es/personal/smp00000/home_archivos/Pay_as_you_drive_directory.htm, 2009.
[48] Directive 2004/52/EC of the European Parliament and of the Council of 29 Apr. 2004 on the interoperability of electronic road toll systems in the Community, Apr. 2004.
[49] Ministry for Urban Development and Roads, "The Controlled Vehicular Access," http://www.cva.gov.mt/, 2007.
[50] D. for Transport, http://www.dft.gov.uk/pgr/roads/introtoroads/roadcongestion/roadpricing demoproject/, 2011.
[51] "Road Pricing: Congestion Pricing, Value Pricing, Toll Roads and Hot Lanes,"technical report, Victoria Transport Policy Inst., http://www.vtpi.org/tdm/tdm35.htm, 2007.
[52] Directive 2006/32/EC of the European Parliament and of the Council of 5 Apr. 2006 on energy end-use efficiency and energy services and repealing Council Directive 93/76/EEC, *Official J. European Union,* vol. 114, no. 27.4, pp. 64-85, Apr. 2006.
[53] M. LeMay, G. Gross, C.A. Gunter, and S. Garg, "Unified Architecture for Large-Scale Attested Metering," *Proc. 40th Ann. Hawaii Int'l Conf. System Sciences (HICSS '07),* p. 115, 2007.
[54] M. LeMay and C.A. Gunter, "Cumulative Attestation Kernels for Embedded Systems," *Proc. 14th European Symp. Research in Computer Security (ESORICS),* M. Backes and P. Ning, eds., pp. 655-670, 2009.
[55] R. Anderson, *Security Engineering.* Wiley, 2001.
[56] A. Escudero-Pascual and I. Hosein, "Questioning Lawful Access to Traffic Data," *Comm. ACM,* vol. 47, no. 3, pp. 77-82, 2004.
[57] E. Barkan, E. Biham, and N. Keller, "Instant Ciphertext-Only Cryptanalysis of GSM Encrypted Communication," *J. Cryptology,* vol. 21, no. 3, pp. 392-429, 2008.
[58] T. Agrelo, "Segurmovil. Automatic Vehicle Tracking," White paper, MAPFRE S.A, 2009.
[59] D. Bell and L. La Padula, *Secure Computer Systems: Math. Foundations and Model.* Mitre, 1974.
[60] A. Menezes, *Handbook of Applied Cryptography.* CRC Press, 1997.
[61] R. Anderson, M. Bond, J. Clulow, and S. Skorobogatov, "Cryptographic Processors—A Survey," *Proc. IEEE,* vol. 94, no. 2, pp. 357-369, Feb. 2006.
[62] D. Naccache and D. M'Raihi, "Cryptographic Smart Cards," *IEEE Micro,* vol. 16, no. 3, pp. 14-24, June 1996.

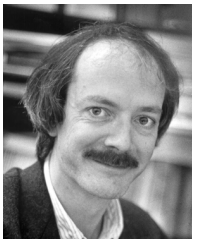
[63] R.J. Anderson, S. Vaudenay, B. Preneel, and K. Nyberg, "The Newton channel," *Proc. First Int'l Workshop Information Hiding*, R.J. Anderson, ed., pp. 151-156, 1996.
[64] G. Simmons, "Subliminal Communication is Easy Using the DSA," *Proc. Workshop Theory and Application of Cryptographic Techniques on Advances in Cryptology (EUROCRYPT '93)*, T. Helleseth, ed., pp. 218-232, 1993.
[65] V.D. Gligor, *A Guide to Understanding Covert Channel Analysis of Trusted Systems*. Nat'l Computer Security Center, 1993.
[66] V. Kirtane and C.P. Rangan, "RSA-TBOS Signcryption with Proxy Re-Encryption," *Proc. Eighth ACM Workshop Digital Rights Management (DRM '08)*, pp. 59-66, 2008.
[67] P. Rogaway and T. Shrimpton, "The SIV Mode of Operation for Deterministic Authenticated-Encryption (Key Wrap) and Misuse-Resistant Nonce-Based Authenticated-Encryption," 2007.
[68] A. Shamir, "How to Share a Secret," *Comm. ACM*, vol. 22, no. 11, pp. 612-613, 1979.
[69] A. Biryukov and E. Kushilevitz, "From Differential Cryptoanalysis to Ciphertext-Only Attacks," *Proc. 18th Ann. Int'l Conf. Advances in Cryptology (CRYPTO)*, pp. 72-88, 1998.
[70] M.G. Kuhn, "An Asymmetric Security Mechanism for Navigation Signals," *Proc. Sixth Int'l Workshop Information Hiding*, pp. 239-252, 2004.
[71] Analog Devices, *3 Axis ADXL330 Low Power Accelerometer Datasheet*, http://www.sparkfun.com/datasheets/Components/ADXL330_0.pdf, 2009.
[72] Siemens, "Anders Betalen Voor Mobiliteit, Phase 2 Market Consultation Report,"http://static.ikregeer.nl/pdf/BLG9687.pdf, 2009.
[73] J. Balasch, I. Verbauwhede, and B. Preneel, "An Embedded Platform for Privacy-Friendly Road Charging Applications," *Proc. Design, Automation and Test in Europe Conf. (DATE '10)*, pp. 867-872, 2010.
[74] ARM, ARM7TDMI Technical Reference Manual, Revision: r4p3, 2009.
[75] NXP, "LPC23xx User Manual," http://www.standardics.nxp.com/support/documents/microcontrollers/pdf/user.manual.lpc23xx.pdf, Apr. 2009.
[76] Keil, "MCB2300 Evaluation Board Family," http://www.keil.com/mcb2300/, Apr. 2009.
[77] Telit, "GM862-GPS Hardware User Guide," http://www.telit.com/module/infopool/download.php?id=871, Apr. 2009.
[78] OpenStreetMap, http://openstreetmap.org/, Apr. 2009.
[79] M. Dworkin, "Recommendation for Block Cipher Modes of Operation: The CCM Mode for Authentication and Confidentiality," Nat'l Inst. of Standards and Technology, NIST Special Publication 800-38C, http://csrc.nist.gov/publications/nistpubs/800-38C/SP800-38C_updated-July20_2007.pdf, 2004.
[80] NIST, *Advanced Encryption Standard (AES) (FIPS PUB 197)*, Nat'l Inst. of Standards and Technology, Nov. 2001.
[81] RSA Laboratories, *PKCS #1 v2.1: RSA Cryptography Standard*, RSA Data Security, Inc., http://www.rsasecurity.com/rsalabs/pkcs/pkcs-1/index.html, June 2002.
[82] A.N. Yannacopoulos, C. Lambrinoudakis, S. Gritzalis, S.Z. Xanthopoulos, and S.K. Katsikas, "Modeling Privacy Insurance Contracts and Their Utilization in Risk Management for ICT Firms," *Proc. 13th European Symp. Research in Computer Security (ESORICS)*, pp. 207-222, 2008.
[83] V. Gratzer and D. Naccache, "Alien versus Quine, the Vanishing Circuit and Other Tales from the Industry's Crypt," *Proc. 25th Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT '06)*, S. Vaudenay, ed., pp. 48-58, 2006.

Carmela Troncoso received the master's degree in telecommunications engineering from the University of Vigo in 2006. Since then, she is a PhD candidate at Katholieke Universiteit Leuven. Her research is focused on Privacy Enhancing Technologies, among them anonymous communications, anonymity metrics, and location privacy. Her research is funded by the Fund for Scientific Research in Flanders (FWO).

George Danezis is a researcher at Microsoft Research, Cambridge. He has been working on privacy enhancing technologies at MSR Cambridge (United Kingdom), K.U. Leuven (Belgium) and the University of Cambridge (United Kingdom), where he completed his doctoral dissertation in 2004. His theoretical contributions to the PET field include the established information-theoretic metric for anonymity and the study of statistical attacks against mix systems. On the practical side, he is one of the lead designers of Mixminion, the next generation remailer, and has worked on the traffic analysis of deployed protocols, such as SSL and Tor. His current research interests focus around peer-to-peer and social network security, as well as the application of machine learning techniques to security problems. He was the cochair of the Privacy Enhancing Technologies Workshop in 2005 and 2006, he serves on the PET Symposium board and regularly participates in program committees of leading conferences in the field of privacy and security.

Eleni Kosta received the law degree and master's degree in public law from the University of Athens in 2002 and 2004, respectively, and the LLM degree in 2005. In the academic year 2004-2005, she followed the Postgraduate Study Programme in Legal Informatics (Rechtsinformatik) of the University of Hanover (EULISP) with a scholarship from the Greek State Scholarships Foundation (IKY). She is working as a legal researcher at the Interdisciplinary Centre for Law & ICT (ICRI) of the K.U. Leuven, where she conducts research in the field of privacy and data protection, specializing on new technologies and electronic communications. She is also working as part-time associate at the law firm time.lex.

Josep Balasch received the master's degree in telecommunication engineering from the Universitat Politecnica de Catalunya, Barcelona, Spain, in 2008. Currently, he is working toward the PhD degree from Katholieke Universiteit Leuven, Belgium, under the supervision of Professor Ingrid Verbauwhede. His research interests include efficient implementations of cryptographic algorithms, embedded security, and privacy-preserving systems.



Bart Preneel received the MS degree in electrical engineering and the PhD degree in applied sciences (cryptology) from the Katholieke Universiteit Leuven, Belgium, in 1987 and 1993, respectively. He is currently full professor with K.U. Leuven. He was visiting professor at five universities in Europe and was a research fellow with the University of California at Berkeley. He has authored and coauthored more than 300 reviewed scientific publications and is the inventor of two patents. His main research interests are cryptography and information security. He is president of the International Association for Cryptologic Research (IACR) and of the Leuven Security Excellence Consortium (L-SEC vzw.), an association of 60 companies and research institutions in the area of e-security. He is a member of the Editorial Board of the *Journal of Cryptology*, the *Journal of Computer Security*, and the *International Journal of Information and Computer Security*. He has participated in more than 25 research projects sponsored by the European Commission, for four of these as project manager. He has been program chair of 12 international conferences (including Eurocrypt 2000, SAC 2005, and ESORICS 2010) and he has been Invited Speaker at more than 50 conferences. In 2003, he received the European Information Security Award in the area of academic research. He also received an honorary Certified Information Security Manager (CISM) designation by the Information Systems Audit and Control Association (ISACA). He is a member of the IEEE.