Privacy Verification Chains for IoT

Noria Foukia[1(✉)], David Billard[2], and Eduardo Solana[3]

[1] University of Applied Sciences, Rue de la Prairie 4, 1202 Geneva, Switzerland
noria.foukia@hesge.ch
[2] University of Applied Sciences, Rue de la Tambourine 17,
1227 Carouge, Switzerland
david.billard@hesge.ch
[3] University of Geneva, Route de Drize 7, 1227 Carouge, Switzerland
eduardo.solana@unige.ch

Abstract. The present paper establishes foundations for implementing Privacy and Security by Design in the scope of the Internet of Things (IoT) by using a new paradigm namely the Privacy Verification Chains (PVC). PVCs will act as a "privacy ledgers" allowing participating entities to prove that they are entitled to hold privacy-related information, regardless of how this information is handled or stored. Furthermore, the PVC structure provides the two following benefits: In case of a security breach resulting in a user data leak, the affected company may browse all the relevant PVCs in order to identify the users affected and trigger the corresponding informative and corrective measures. The PVC will also provide support for bidirectional browsing which means that the data owner will be capable of browsing all the PVCs involving the data he owns in order to find out all the data processors that hold his personal information. From a wider perspective, we enforce a strict separation between data providers and data controllers, where providers are managers of their data privacy, and controllers are accountable for the privacy and protection of the data provided. This role separation will be ensured by a data controller of a so-called Smart Data System (SDS). The SDS handles information along with its privacy settings (metadata), defined by the data owner. In order to control this privacy-preserving framework, our system introduces a Forensic and Auditing System that will enforce the data protection from the processor to a third party. This component will also provide a comprehensive logging functionality that will constitute a legally-binding support to respond to audit procedures, police investigations and(or) law enforcement obligations.

Keywords: Privacy by Design · Internet of Things · Privacy Verification Chains

1 Introduction and Context

Since the very beginning of the digital era, one of the most significant challenges has been the protection of data against non-authorized reproduction, redistribution and use. Both the academic and the private sectors have pervasively

Z. Yan et al. (Eds.): NSS 2017, LNCS 10394, pp. 737–752, 2017.
DOI: 10.1007/978-3-319-64701-2_58

addressed this issue for years with a certain level of success. Numerous research and commercial products have resulted in workable solutions for data embedded in specific structures (such as images, worksheets, processed texts, etc.) and physical supports (DVDs, hard disks, USB disks, etc.). Watermarking and Digital Rights Management techniques have paved the way to detect fraudulent activities related to these categories of structured information. In most cases, a cryptographically computed combination of identifiers, digital signatures and data digests is smartly embedded in the original data in such a way that extraction and/or modification of this watermark is computationally hard without modifying the data to protect. Whereas these solutions have been widely implemented and are regularly used to protect copyrights of images, disks, logos, etc., none of the existing initiatives has successfully addressed the issue of protecting non-structured information such as textual data strings, telephone numbers, e-mail addresses, in which no watermark can possibly be inserted. As a result, an enormous amount of these non-structured data is freely used and managed by states, companies and individuals with little or no control by data owners.

Regulators and governments inside and outside the European Union (EU) have consistently tried to implement legal frameworks to protect the processing and the free movement of personal data within the EU [1–3]. Unfortunately, in practice, this legal framework has not produced the intended results and the actual level of citizen privacy protection is well below any reasonable standard.

Although personal information is often collected through a set of processes supposedly compliant with national or international privacy laws, the effective way this information is used (and misused), managed, distributed and even sold to third parties is well beyond the actual data owner control. There are a variety of reasons behind this difficulty to implement an effective privacy protection policy but in our opinion the most important one is the inherent facility to copy and distribute digital information especially since the inception of the Internet. A worksheet containing millions of records with user personal information can be exchanged in a fraction of a second, seamlessly and without leaving any consistent trace.

Many companies grant personal information a huge economic interest using it to profile consumers' behaviour in order to adapt their advertising strategy. Whether in the EU, in Switzerland or certainly all over the world, most people ignore this massive personal information collection and usage. For instance, in the Swiss Federal Data Protection Law [3], the right for each citizen to exercise full control of his personally identifiable data is fundamental. This control rules the way information is transmitted, managed and may be redistributed by involved third parties. Besides, legitimate interests may limit this right (e.g., fight against crime, related police investigations or other empowered authorities). In this case, the proportionality principle must be respected meaning that the data collection and processing should involve as little personal data as possible and not more than is absolutely necessary [3]; this also means that the data subject can check at any time the processing of his own data and, if necessary, object to it.

Furthermore, the emergence of new gadgets (smart wearable objects) and technologies has provoked an evolution on our society in which a need for being connected and exposed to others has appeared. We have gone from exchanging emails to share our position (outside and inside buildings), vital signs or social interactions making impossible to control how and who is accessing that data. Therefore, the relatively new Privacy by Design (PbD) [18] paradigm tends to evolve from essential to mandatory in the scope of EU data protection directives [16,20], meaning that data protection safeguards should be built into products and services from the earliest stage of their conception.

In regard to this PbD recommendation, the present paper aims at establishing foundations for implementing Privacy and Security by Design (PSD) in the scope of the Internet of Things (IoT) by using a new paradigm namely the Privacy Verification Chains (PVC). Concretely, we take the perspective to operate a strict separation between data providers and data controllers, where providers are managers of their data privacy, and controllers are accountable for the privacy and protection of the data provided. This role separation will be ensured by a data Controller of a so-called Smart Data System (SDS). This data controller handles data along with its privacy settings (metadata), defined by the user.

Allowing users to take care of their privacy, while respecting state requirements (w.r.t. law enforcement) and freedom of business must be considered as a fundamental priority. Thus, the SDS allows balancing user privacy against the need to access information in case of law-enforcement organization activities (e.g., police investigations in fight against crime) or other legitimate activities (e.g., patient health service survey). This is made possible thanks to the PVC allowing the data owner and/or any intermediary (data controllers, data processors) to know easily by whom, and for which purpose, the data is used, thus asserting whether the users' rights are respected or not. From an economic perspective, the system enables Internet users and service providers to get a reasonable bargain when monetizing user data; it makes a necessity to define fair and mutually acceptable conditions for using the services and the data. These conditions can give incentives for the user to grant access to his data and for the service provider to facilitate free usage of some services (e.g., value-added information channels, free access to email or social platforms, etc.).

We propose to implement privacy by reversing the way we currently look at it. Today:

- Privacy and security are managed by the application, i.e. each application imposes its own schema for privacy and security, often very complex to handle, and non-interoperable.
- The security of data relies on the security of the OS and applications (a breach of the OS or a badly implemented application means a disclosure of all the data).
- Copies of redundant data are held by several applications and entities, multiplying the risk of disclosure.
- The user totally ignores who is using his data and to which purpose.
- The law enforcement entity has to deal with scattered and low quality data in investigations.

Thus, our objective is to provide:

- A separation of the data provider and data controller, especially for the IoT, in order to prevent devices accessing to the Internet, connecting and sharing data with other products, without the informed decision and control of the user.
- An increase in data security by the design of a privacy system, namely the SDS.
- A simple way for the Internet user, through the PVC use, to exercise full control of his personally identifiable information including entities authorized by him to hold or/and manage this data.
- A smart mechanism for the Internet user, again through the PVC, to understand the source and extent of a data violation (breach or misuse).
- A simple means for the Internet user to set the privacy level he intends for his data through the SDS.
- A simplified process for companies and users to reach a fair bargain on the usage of private data.
- A well-defined mechanism for law enforcement to access critical information.
- A simple way for independent administrative authorities operating data protection legislation to effectively control the usage of data.

In Sect. 2, we summarize the state of the art in the field of blockchain principles, IoT operating systems, IoT forensics and Privacy by Design. Section 3 describes the SDS model principles including the PVC. Section 4 provides the SDS architecture. The conclusion is given in Sect. 5.

2 State of the Art

2.1 Blockchains

The issue of privacy preserving using a distributed peer-to-peer model has been recently addressed in [5] where the authors explore the use of blockchain technology to protect privacy. They introduce an access control manager where the blockchain holds privacy preserving policies involving users (data owners) and services (data processors). Due to the public nature of blockchain transactions, sensitive data is stored in an off-blockchain structure based on Kademilia, a Distributed Hashtable (DHT) maintained by a network of nodes as described in [6]. The authors acknowledge the issue where services may extract raw data from the DHT and make unlimited use of it so they propose a secure Multiparty Computation (MPC) to evaluate functions. The range of operations that can be effectively achieved by this means remains quite limited. Beyond blockchains, fully homomorphic encryption as described by Craig Gentry in his seminal paper [7] constitutes the most promising research direction to generically address privacy protection in the encrypted domain. Unfortunately, practical implementations of this groundbreaking work remain inefficient by the time of writing.

2.2 Internet of Things OS and Forensics

Our society is witnessing a "rush to market" by multiple vendors to launch a wide variety of IoT devices addressing each and every aspect of human's life. Unfortunately, no consensus has been reached so far on how privacy-sensitive data is collected, handled or even monetized. This fact is highlighted in [8] where authors demonstrate that there is no established best practice for the building of IoT systems. The vast majority of products uses proprietary software, or rely on an open-source framework, like Busybox, a lightened version of Linux, whose slogan is: "The Swiss Army Knife of Embedded Linux".

In [9], the authors present the advances in low-power protocols (like 6LoWPAN or CoAP) and sensor nodes and propose a software architecture to handle IoT. However, the authors focus on the application level and rely on a traditional operating system, thus missing the point of security at a low level. A survey of some manufacturers proposing proprietary IoT platforms can be found in [10]. This very comprehensive work shows the diversity of the market and classify the IoT depending on their ability (or inability) to perform: (1) context aware tagging and (2) context selection and presentation. The authors further present the challenges of prototyping IoT software, when do-it-yourself (DIY) prototyping is the rule by using Arduino (https://www.arduino.cc/), Raspberry Pi (https://www.raspberrypi.org/), .NET micro platform (http://www.netmf.com/) or LittleBits (http://littlebits.cc/). Finally, [11] provides an analysis of IoT threat models, security issues and forensics requirements. The authors define five major components in an IoT ecosystem: (1) devices, (2) coordinator (device manager), (3) sensor bridge (or gateway to the services in the cloud), (4) services (cloud-based applications) and (5) controller (the user accessing the services via smartphones or computers). This definition of an ecosystem is the closest to reality. Furthermore, the authors list security constraints and requirements at several levels, like hardware (they advocate for tamper resistant packaging) or software (thin, robust and fault-tolerant security module). In their work, the authors deliver a comprehensive and accurate landscape for IoT security, but do not present any solution to the security challenges. For instance, concerning the IoT forensics, the concluding remark is: "The definition of an efficient and exact IoT digital forensics procedure is still at its great demand". In [12] the authors draw the attention of the research community about the lack of digital forensics devices, specific applications or even guidelines to support potential IoT investigations. This lack of any serious research in IoT forensics (to the extent of our literature review) is persistent through all the literature. To the best of our knowledge, [13] constitutes the first and only effort to model the capture of forensics investigations when IoT devices are involved. The authors define three schemes for forensics investigation: (1) device level forensics, (2) network forensics and (3) cloud forensics. This model, called FAIoT, is a first attempt at defining IoT forensics and, as such, it should be considered as a seminal work in the field. However, we postulate that IoT forensics is far more than a juxtaposition of known forensics venues. The model concentrates on the hacking of IoT devices whereas all the area of identifying data leakage and tampering is left aside. It focuses on the traces left by an attacker, leaving aside the investigation of the data flows. Understanding the nature of an attack upon IoT is, of course, essential, but in order to be complete and to follow forensics principles, we should be in position to track the data flows. This model reflects the traditional foreseen usage of IoT in criminal cases: hacking to take control of the devices. For instance, hacking an insulin pump in order to blackmail a user, or taking control of Supervisory Control And Data Acquisition (SCADA) systems. These constitute extreme cases, that need to be addressed, but we consider that the most representative ones relate to data breaches and illicit private information disclosure or tampering. Most frequently, IoT devices constitute the origin or the main propagation vector of these incidents.

As a matter of fact, among the top cybersecurity threats and cases documented in the scientific literature throne data breaches [14,15]. At Swiss and EU level, a very strong move towards data protection is operated, and the need of forensic evidences is in high demand.

2.3 Privacy by Design

In addition to ongoing regulatory effort in the EU [1,2] and in Switzerland [3], the PbD principle appeared, mainly cultivated by Cavoukian in Canada since 2008 [17]. According to Cavoukian, PbD is based on seven foundational principles, although new additions to this list, such as the data minimization principle have recently emerged. In these principles, Cavoukian [17], pointed out the fact that the legal framework was not sufficient to ensure the protection of the private sphere. In fact, the European Union Data Protection Directives [1,2] always require data controllers to implement appropriate technical and organizational measures for personal data protection. However, this has proven to be insufficient since often the data protection and privacy principles are added as an additional layer over existing ICT systems.

In order to remedy this technical weakness in terms of privacy protection of ICT systems, Ann Cavoukian proposed to directly integrate preserving privacy means from the start of the system design and during the system operation. She encouraged the usage of Privacy Enhanced Technologies (PETs) when possible and she also urged the usage of Privacy by Default (see below the second principle proposed by Cavoukian). Ideally, as follows from the definition of the European Parliament commission [19], a PET should act as a coherent system of ICT measures that protects privacy by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data, all without losing the functionality of the information system. More precisely, PbD means that the requirements in terms of privacy protection and data protection will be taken into account at the earliest stage of the product design. Concretely, from the work done by Cavoukian et al. [17], the seven key PbD recommendations that emerged are listed below:

1. Proactivity rather than reactivity: privacy measures need to be taken before the privacy-invasive events happen (prevention and minimization).

2. Privacy by Default: default settings need to ensure the maximum degree of privacy and data protection without direction from the data user.
3. Privacy Embedded into Design as an essential component integrated in the whole ICT core system.
4. Full sum (no trade-off): ICT systems need to include privacy from the start without making any un-relevant trade-off such as increasing security to the detriment of privacy.
5. Full life cycle: Privacy included into the ICT system design from the start: before any data have been collected in the system, during the entire system operation and also during the entire life cycle of the data.
6. Visibility and Transparency: At any moment, the data user should be given the possibility to know and control who has his data, what data have been collected and for what purposes they will be used in accordance with the legitimate initial purpose.
7. User centric: Privacy by default measures (such as opt-in option), appropriate notice for privacy settings selection should enable the data owner to quickly and easily obtain the highest level of protection.

Recently, in May 2016, the official texts of the new EU Regulation and the Directive have been published in the EU Official Journal [20]. The new GDPR was published on May the 4th, 2016 although enforcement will be effective on May the 25th, 2018. The regulation focuses on "a consistent and high level of protection of the personal data of natural persons but also facilitates the exchange of personal data between competent authorities of Members States. This is crucial in order to ensure effective judicial cooperation in criminal matters and police cooperation." The recent GDRP regulation [16, 20] also puts forward the PbD paradigm as a fundamental principle for implementing privacy controls at all the different stages of the information life cycle. In particular, we consider especially relevant the Article 20 titled "Data protection by design and by default" that clearly states EU regulators' recommendation to "implement appropriate technical and organizational measures (such as pseudonymization), which are designed to implement data protection principles, such as data minimization, in an effective manner and to integrate the necessary safeguards into the processing, in order to meet the requirements of this Directive and protect the rights of data subjects". However, no indication is provided on how to accomplish these ambitious and complex objectives and, as a consequence, we consider that our work constitutes a significant contribution to fulfill this goal.

3 The Proposed Model - Smart Data System

3.1 Setting the Privacy Level of Data

Our conception of the data system is based on the primary providers of data: the IoT/Internet users. They can define the level of privacy for their own data and attach requirements to their data. Since we address the average Internet users, the system should make the operation simple to understand and easy to realize. Therefore, the privacy level will be universal and easily understandable; it comes into three privacy levels: Public, Confidential, Intimate. Each level should be in accordance with the EU data protection regulations or national privacy protection laws.

In terms of ergonomics and ease of understanding by all kinds of users (who can also include elderly or disabled people), we propose a graphic scale helping the user to express the sensitivity of the provided data. This scale might be similar to the energy consumption scale that can be found for different home appliances. An example is provided in Fig. 1: The privacy level will be linked to the intended usage of the data. For instance, age and weight of a person can be public for an anonymous statistical review, and confidential, i.e. restricted, to the application managing the user's health.

3.2 Controlling Data Ownership and Usage

The right to use the data is transmitted from the provider to the controller. The data controller is granted the right to use the data, but never the ownership of the data. The usage of the data can be multiform, for instance:

- Statistical and anonymous usage of data;
- Business usage for commerce and marketing;
- Intelligence usage for other consumers, producers and others.

Furthermore, the data is transmitted with metadata describing the data ownership, its transmission mode and security tokens, through the PVC. A contract links the data controller and the data provider upon the delivery of data and its usage. Financial agreement can be reached. All this meta information should be associated to the data.

3.3 Role of a Smart Data System

We propose a component called Smart Data System (SDS) to handle the data privacy level, the data intended usage and the metadata. The SDS is designed in a way to achieve a strict segregation between the data provider and the data controller, preventing the former from communicating directly through the Internet. By preventing direct communication, we position the user as a central actor who can control and qualify his data, in a simple and seamless way via the SDS. This implies that only the SDS is allowed to communicate with the Internet. The data provider (the user himself or an appliance) may only communicate with the SDS. The SDS is designed with the following assumptions:

- Failures will arise
- Security attacks will be launched
- Data breaches will happen
- Data corruption will occur

Hence, the SDS will have to implement mechanisms to handle the above enumerated problems that are the consequence of a normal course of events in the present Internet. Experience shows that failures, attacks and/or breaches are not an exception but rather a rule in the operation of the global Net.

3.4 SDS Architecture

We propose a division of SDS into several containers (sand boxes) monitored by fault-tolerance and intrusion detection services. These containers have the capability to be easily distributed to cope with heavy load and moveable to cope with attacks. For instance, the SDS can provide an anomaly detector, a fault handler and can issue service notification messages such as: Security Breach, too much data requested, an improper use of data, etc. If an unusually huge amount of data or malformed data is sent to the SDS, the anomaly detection system should detect it and send a signal to other containers so that they can stop communicating and shut down. A similar response may be triggered if a container is no more answering a probe signal. The SDS should be independent from any OS and/or application, so that it can be placed either in a cloud, a home computer, or a small device such as Google glasses.

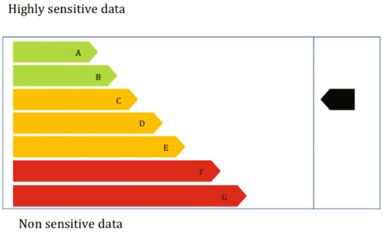

Fig. 1. Data sensitivity scale

3.5 Ensuring Fight Against Crime

User privacy should be balanced against the need to access the information in case of police investigations (fight against crime). For this purpose, at the provider side, the SDS will ensure safe logging of the transferred data and will guarantee its intended usage. In case of consumer abuse, the forensic logs can be provided to a court. The data traceability and accountability are implemented by a separated service running side by side with the SDS. At the consumer side, traceability and accountability of data is implemented, as required by each country laws.

We are considering homomorphic cryptography a strong candidate to safeguard and present the evidences. In fact, the homomorphic cryptography would link together different services acting on the data without exposing the data to each of those services. Thus, with homomorphic encryption, the use of cloud services for forensic soundness is feasible. For instance, new services acting as trusted third party, can store forensic logs and eventually provide them upon legal warrants.

3.6 Monetization of Data

Currently, while most of the Internet users are giving their data for free, the majority of Internet service providers are using this data for financial gains as part of their business models. In order for Internet users and service providers to get a reasonable bargain, it is necessary to define fair and mutually acceptable conditions for using the services and the data. In addition, these conditions can be dealt with in such a way that they give incentives for the user to allow more access to his data and for the service provider to allow free usage to some services. For instance, a service can be provided for free (i.e. without currency being exchanged) if the user agrees to give the full usage of its private data to the provider. If the user is not willing to give the provider the usage of its data, then the provider can charge the user for its service. Otherwise, if the user considers his data are very valuable, he can charge the provider for the usage of his data.

3.7 PVChains Scenario

Though the SDS design is not limited to IoT, it is expected that some modification will be needed for any other type of data. The IoT scenario includes a Digital Weighting Scale, a common appliance. The device registers to the SDS and sends its data. This part (the registering and the communication) is not designed. The very important data of 90 Kg and 80 pulses per min is sent to the SDS, which stores it, along with the metadata, inside a data repository. The SDS publishes the data tagged Health on the channel Confidential. The Internet application for the Weighting Scale, also registered with the SDS, can retrieve only and exclusively this data.

4 Detailed SDS Architecture

Figure 2 provides a detailed view of the SDS functional architecture.

4.1 The Controlling Area

The controlling area is inside the CSDS (Controller Smart Data System). The purpose of this area is to offer private data management for the IoT. It is composed of:

- DataManagementFromIoT module
  - When the user is buying an IoT device, the device is not able to communicate to the Internet. The user has to configure it via a network connection. This connection, usually initiated via a web interface (the device running a small web server for administration purpose), will now be initiated through the DataManagementFromIoT module. - The DataManagementFromIoT exchanges information with the device: public keys, name to be displayed on the device screen (if any), etc. - The device sends the URL of the data processor web site. - The device sends the complete list of data

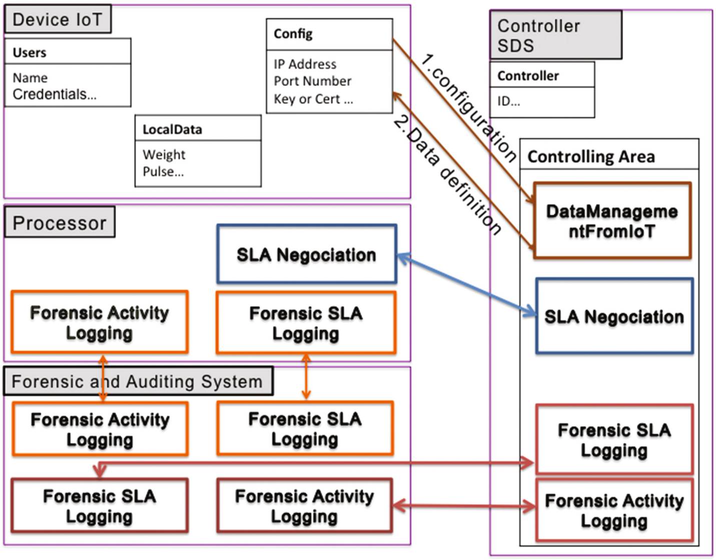

Fig. 2. SDS architecture

definition that it will collect. For instance, for a scaling machine: Date and Time (this is the date of the measure, YYYYMMDD-hh:mm:ss), weight (this is your weight in Kg), heartbeats (this is your pulse rate per second), fat (this is your fat percentage) and user (this is the user using the scale). - The DataManagementFromIoT presents all these data definitions to the user who is required to choose a privacy level for each of them. - The DataManagementFromIoT connects to the processor URL and sends the list of data definition and privacy settings. In this first step, a minimal load has been put on the device, which often has limited memory or/and processing power. It also imposes minimal requirements on the IoT provider with respect to possible modification of the software in the device. Once this first step is over, the data controller and the processor have to reach an agreement on the data and the data usage. This agreement is drafted in the form of a Service Level Agreement (SLA).

– SLA negotiation
  The pricing of the data is settled. The negotiation of the price could be done in terms of money or service. Once the SLA is ready, the controller and the processor send their agreement to the Forensic and Auditing Facility, which is an external service. The intent of agreeing is logged, and the reached agreement is logged too.

4.2 The Forensic and Auditing System

The Forensic and Auditing System (FAS) is the component that mostly enforces the data protection part of the proposal, whereas the Controller Smart Data System defines the data privacy part. The FAS has several missions:

- Logging the establishment of an SLA between a controller and a processor. The transaction is done by recording both IDs, SLA (content or digest, tbd), timestamp of issuance and Time to Live (TTL, or duration); the SLA itself might be a list of data definitions, properties or use cases. This logging is done by the ForensicSLALogger.
- Logging the data transmission from the controller to the processor. The data itself do not need to be logged, but at least the IDs, data requested, data granted, timestamp of demand and TTL. The logging is done by the ForensicActivityLogger.
- Logging the data transmission from the processor to a third party. This activity is important. It enables the processor to document where/when the data is transferred, and in the same time, the third party can verify at the FAS if the processor has the right to sell the data, according to the SLA between controller and processor. The third party in turn becomes a processor.
- Answering police, or audit, investigations by providing the records upon justice warrant. Logging systems have been extensively used in operating systems, and more particularly Database Management Systems, in order to provide transactional services (Atomicity and Durability). In SDS, the logging system is designed for the safekeeping of three kinds of evidence, should a dispute arise:
- Evidence of an SLA establishment, and the SLA content.
- Evidence of data served from the controller to the processor.
- Evidence of data transmission from the processor to a third party.

It is a forensically oriented logging system. Note that a traditional logging system also exists for transactional services inside the SDS. In order to maintain the quality of the stored evidence, the ForensicActivityLogger and the ForensicSLALogger must be duplicated at the controller side and processor side. We propose to use homomorphic cryptography to ensure a secure environment. The homomorphic cryptography permits to process forensic logs under the same secret at the different location, without an extensive key management.

4.3 Privacy Verification Chains

We propose a system that links private information to a control structure that provides a proof of legitimacy that can be validated by regulators and police authorities. This control structure is based on a Peer-to-Peer (P2P) paradigm where the Data Owner (provider) and the Data Processor would digitally sign (mutual signature) a Privacy Control Record (PCR) containing amongst others the following entries:

- Data owner ID and Data Processor ID. Both digital identities as they appear in the public key certificate used to validate their signatures.
- The Data Description field of the information subject to the contract: for instance, "name", "e-mail address", "height", "weight", "location", "beats per minute", etc. It should be noted that only the header and not the data itself would be included in the PCR.
- The SLA indicating the actions that the processor may or may not achieve on the relevant information. This may include: storage, encrypted storage, limited/unlimited distribution, selling to a third specific/non-specific third party, etc.
- Contract date, Time to Live, Contract Expiration, Data Expiration, etc.

As an example, if the SLA allows the Data Processor 1 to distribute the data to a so-called Data Processor 2, this operation would result in a new PCR that would consist in a signed (by Data Processors 1 and 2) record containing the same entries plus a pointer to the previously described PCR signed by the Data Owner and Data Processor 1. This new PCR would include a new SLA that depends on the terms of the first agreement in such a way that if the new SLA authorized the distribution to a given entity other than Data Processor 2, the new PCR would be invalid. The resulting control structure would be a chain of PCRs that we name the Privacy Verification Chain (PVC) where the first PCR would be the one signed by the Data Owner and Data Processor 1 (Fig. 3). Given this control structure, how can a given organization prove that it is entitled to hold personal information related to a user? By providing to the law enforcement processor a pointer to a PVC which links would be iteratively validated (organization/Law Enforcement browsing mode in Fig. 3.) up to the first one containing the data owner whose personal information is affected. In other words, the PVC acts as a Privacy Ledger that allows participating entities to prove that they are entitled to hold privacy-related information, regardless of how this information is handled or stored. Furthermore, the PVC structure provides other relevant uses and benefits:

- In case of a security breach resulting in a user data leak, the affected company may browse (organization/Law Enforcement browsing mode in Fig. 3) all the relevant PVCs in order to identify the users affected and trigger the corresponding informative and corrective measures. This feature provides significant benefits to companies complying with the new EU GDPR regulations [20] where the maximum notification delay for a breach affecting, so called, Personally Identifiable Information is limited to 72 h.
- The PVC will also provide support for bidirectional browsing which means that the data owner will be capable of browsing all the PVCs involving the data he owns (data owner browsing mode in Fig. 3). This represents a relevant achievement since it allows the user to discover all the data processors that hold his personal information, including these entities that have not established direct agreements with him.

As explained, the control structure does not hold personal user information and, as a result, does not need to be kept confidential. Only authentication and

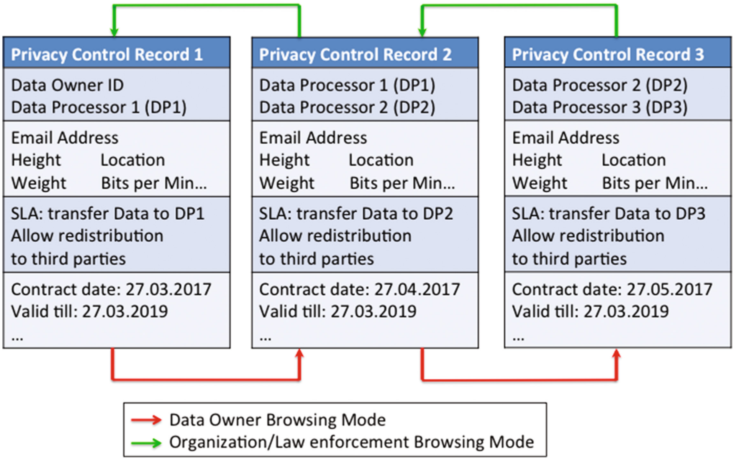

Fig. 3. Privacy Verification Chain

signature validation are required for every PCR link in the PVC. If the identities of the signing parties are considered sensitive, PCR records may be represented as cryptographic digests facilitating verification without disclosing PCR contents. Regarding the confidentiality of the personal information, it should be noted that the SLA may include policies stating how information should be kept and managed at the processor facilities. For instance, if the signed agreement enforces that personal information should always be stored encrypted and the processor keeps this data on the clear, it may be subject to fines or sanctions from a law enforcement entity.

5 Conclusion

This work provides an increase in data security by the design of a privacy system, the SDS. The SDS furnishes a simple way for the Internet user to set the privacy level he intends for his data. It also provides an intuitive mechanism for the Internet user, through the use of PVC to know who is using his data, which data, and to what purpose. This innovative Privacy Control Infrastructure based on a peer-to-peer model allows law enforcement entities to find out whether data controllers are entitled to hold personal user information. Moreover, PVChains may be considered as (1) a structured and organized framework for companies and users to reach a fair bargain on the usage of private data (2) a set of well defined procedures and controls for law enforcement to access critical information (3) simplified process for independent administrative authorities operating data protection legislation to effectively control the usage of data.

References

1. EU Data Protection Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal L281, pp. 31–50, 23 Nov 1995
2. EU Directive 2016/680 the European Parliament and of the Council, Official Journal, 27 Apr 2016
3. Confédération Suisse, Avant-projet de la Loi fédérale sur la protection des données (LPD)
4. Foukia, N., Billard, D., Solana, E.: A Framework for Privacy by Design in IoT, presented at the Privacy, Security and Trust Conference, Auckland, New-Zealand (2016)
5. Zyskind, G., Nathan, O.: Decentralizing privacy: using blockchain to protect personal data. In: Security and Privacy Workshops (SPW), IEEE, pp. 180–184 (2015)
6. Maymounkov, P., Mazieres, D.: Kademlia: a peer-to-peer information system based on the xor metric. In: International Workshop on Peer-to-Peer Systems, pp. 53–65 (2002)
7. Gentry, C.: Fully homomorphic encryption using ideal lattices. In: STOC, vol. 9, pp. 169–178 (2009)
8. Zanella, A., Bui, N., Castellani, A., Vangelista, L., Zorzi, M.: (2014) Internet of things for smart cities. IEEE Internet Things J. **1**(1), 22–32 (2014)
9. Mainetti, L., Mighali, V., Patrono, L.: A software architecture enabling the web of things. IEEE Internet Things J. **2**(6), 445–454 (2015)
10. Perera, C., Liu, C.-H., Jayawardena, S.: The emerging internet of things marketplace from an industrial perspective: a survey. IEEE Trans. Emerg. Top. Comput. **3**(4), 585–598 (2015)
11. Hossain, M.-M., Fotouhi, M., Hasan, R.: Towards an analysis of security issues, challenges, and open problems in the internet of things. In: IEEE World Congress on services (SERVICES), pp. 21–28 (2015)
12. Watson, S., Dehghantanha, A.: Digital forensics: the missing piece of the Internet of Things promise. Comput. Fraud Secur. **2016**(6), 5–8 (2016)
13. Zawoad, S., Hasan, R.: FAIoT: Towards building a forensics aware eco system for the internet of things. In: IEEE International Conference on Services Computing, pp. 279–284 (2015)
14. Liu, Y., et al.: Cloudy with a chance of breach: forecasting cyber security incidents. In: USENIX Security, pp. 1009–1024 (2015)
15. Verizon 2016 Data Breach Investigations Report (2016)
16. European Parliament, European Parliament Legislative Resolution of 12 on the Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation), (COM(2012) 0011 C7–0025/2012 2012/0011(COD))
17. Cavoukian, A.: Privacy by Design - The 7 Foundational Principles, originally published on August 2009, revised on January 2011. https://www.ipc.on.ca/wpcontent/uploads/Resources/7foundationalprinciples.pdf
18. Cavoukian, A.: Operationalizing Privacy by Design: A Guide to Implementing Strong Privacy Practices, December 2012. http://www.cil.cnrs.fr/CIL/IMG/pdf/operationalizing-pbd-guide.pdf

19. Borking, J.: Organizational adoption of privacy enhancing technologies (PET). In: Computers, Privacy and Data Protection: An Element of Choice. Springer, Netherlands, pp. 309–341 (2011)
20. EU Directive 2016/680 the European Parliament and of the Council of 27 April 2016. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L:2016:119:01:0089:01:ENG