Towards privacy and ownership preserving of outsourced health data in IoT-cloud context
Yousef Ould-Yahia, Samia Bouzefrane, Hanifa Boucheneb

To cite this version:
Yousef Ould-Yahia, Samia Bouzefrane, Hanifa Boucheneb. Towards privacy and ownership preserving of outsourced health data in IoT-cloud context. 13th International Symposium on Programming and Systems, Apr 2018, Alger, Algeria. 10.1109/ISPS.2018.8379018 . hal-02922700

HAL Id: hal-02922700
https://hal.science/hal-02922700
Submitted on 26 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Towards privacy and ownership preserving of outsourced health data in IoT-cloud context

1st Youcef Ould-Yahia  
CEDRIC Lab.  
CNAM  
Paris, France  
youcef.ouldyahia.auditeur@lecnam.net

2nd Samia Bouzefrane  
CEDRIC Lab.  
CNAM  
Paris, France  
samia.bouzefrane@lecnam.net

3rd Hanifa Boucheneb  
CEDRIC Lab.  
CNAM  
Paris, France  
hanifa.boucheneb@polymtl.ca

Abstract—In this paper, we propose a novel data-owner centric privacy model for in-home-monitoring applications, that implements a promising attribute-based encryption (ABE) to reinforce the data-owner access control and the security of anonymous data access. This proposed protocol avoids threats from curious cloud service providers. Unlike other schemes that implement ABE by outsourcing the heavy computational tasks such as encryption and decryption processes, we propose a framework in which we externalize the complete ABE-encryption algorithms to avoid complex outsourcing process and use well know efficient symmetric encryption in constrained devices. We have performed an experimental analysis to show how much gain allows such offloading.

Index Terms—e-Health, IoT, Cloud, Privacy, ABE.

I. INTRODUCTION

Privacy has become one of the main concerns in healthcare applications from the perspective of data ownership. This concern is exacerbated by the rapid development of Internet of Things (IoT), where connected devices collect a substantial amount of personal data, such as locations and physiological parameters, and transmits them to the cloud in order to store, process and share them, thanks to information and communication technologies. When health science is combined with IT, the result is the e-Health paradigm. Unlike telemedicine, e-Health is not exclusively reserved for healthcare professionals. E-Health model is rather centered and pivoted on the consumers of health systems [1]. But one critical issue of the e-Health paradigm is that it deals with personal sensitive data that can harm the data privacy. Furthermore, some threats may arise from service providers since many companies have significant commercial interests in collecting private health data [2]. There are also other risks to provide her/his personal data to the service providers that are not completely free from security breaches [3]. Additionally, even the so called trusted third party may deliberately or by inadvertence perform illegal activities on personal healthcare data. In this paper, we consider that privacy protection includes data-ownership control of encryption and access [4], and avoids leakage of information that can be deduced from available metadata (informations on data: date, time, kind of measurement, etc.) [5]. To meet these requirements, we propose a new security model implementing a recent and promising cryptographic scheme called attribute-based encryption (ABE). Our solution is a realistic proposal for in-home monitoring applications that involve largely IoT devices. The main objective of this paper is the data and privacy protection while giving to the data owner a better control on his/her data encryption and sharing.

The remainder of this paper is organized as follows: in Section 2, a brief description is given regarding the context and the environment of in-home monitoring, the challenges and threats on privacy, the regulation and the associated legal constraints. Section 3 is devoted to the literature review on data privacy for e-Health in IoT-cloud environment. It also presents the principle of attribute-based encryption mechanism. In Section 4, our proposed model is described followed by a security analysis, while a formal validation and an experimental analysis to validate the concept are presented in Section 5. Finally, we conclude the paper with some perspectives.

II. IN-HOME MONITORING: CASE DESCRIPTION AND THREAT MODEL

One interesting scenario in e-Health is in-home monitoring as it brings an IoT-cloud architecture and strong privacy constraints.

a) In-home monitoring case description: In-home monitoring technology includes many sensors and smart devices that collect data related to health parameters like blood pressure, heartbeat, etc. to be transmitted for processing, storage and sharing within the cloud. In this context, many commercial solutions are being developed to assist the patient to deliberately share his own health data with health specialists [4]. However, as we can see below, the cloud cannot be considered as completely trustful.

b) Threat model: To define an adversary model, the in-home monitoring in the real world should be considered. The purpose of our proposed solution is to retain and to preserve the privacy of the patient. Therefore, threats derived from the communication channel must be considered. There is an adequate overview provided by the Dolev-Yao model [6], which assumes the network as an intruder. Thus, an attacker can listen, delete, replay and modify a message. However, the attacker will not be able to decrypt a message, if he does not possess the decryption key. Then, since cloud infrastructures are under control of a third party that may be curious [3], we adopt the same model as in [7], that considers
the cloud as a honest but a curious actor. This means that the cloud rolls out the protocol correctly but could try to deduce information about the patient. Furthermore, this model supports the realistic threat involved in the cloud context where a cloud infrastructure can be targeted by malicious entities or can have inadvertent leak personal information.

III. REVIEW OF THE LITERATURE ON DATA PROTECTION AND PRIVACY

a) Privacy in IoT-cloud: Integrating IoT with cloud computing provides a promising solution for managing healthcare sensor data efficiently [8]. However IoT-cloud is a heterogeneous environment. In addition, offering an end-to-end solution in that environment to protect data and privacy is difficult. Hence, the cloud is a resourceful entity that provides unlimited virtual resources while its security and trustworthy are managed by a third party [3]. Whereas, IoT may have a high level of trust with a physical control on the devices while facing resource restrictions. Relevant literature reviews bring out that the solutions offered are confronted with this issue. For example, authors of [3] and [7] provide an interesting solution to achieve a fine-grained access control on cloud environment, but did not deal with restrictive IoT environment. However in [9] and [10], the authors propose to secure data and protect privacy in IoT environment without evidence of privacy in the cloud side. The approach proposed in [11], [12] and [13] is based on Attribute-Based Encryption (ABE), by outsourcing heavy computational operations. This scheme provides a patient-centric encryption and access control. However, the proposed solution needs a significant increase in messages exchanges, furthermore [11] and [12] did not propose to outsource the encryption process,which is needed in some IoT-health applications.

b) Attribute-Based Encryption (ABE): Recently, the promising ABE scheme is proposed to enhance privacy for e-Health applications like in-home monitoring [14]. ABE is a public key one-to-many encryption scheme. In addition to securing data transmission and storage, ABE provides a fine-grained access control, a scalable key management and a flexible data distribution [15] and [7]. It allows encrypting the data without any prior knowledge of the identities of the recipients. Two principal variants are proposed: Ciphertext Policy Attribute Based Encryption (CP-ABE) [16] and Key Policy Attribute Based Encryption (KP-ABE) [17]. We focus on CP-ABE (Fig.1) which allows the data owner to encrypt and define access policy for his data. The algorithms associated to CP-ABE are reported in table I. ABE is a pairing based encryption scheme, using a bilinear map denoted \( e : G_0 \times G_0 \rightarrow G_T \), where \( G_0 \) and \( G_T \) are a bilinear group of prime order \( p \). The usual implementation of ABE builds a group from an elliptic curve \( E \) defined over a finite field \( F_q \), to obtain a multiplicative cyclic group of prime order \( p \), where the field size \( q \) and the prime order \( p \) determine the security strength [15]. For a security level corresponding to AES 128 bits, as shown in table II, we choose an elliptic curve defined over a finite field \( F_{2^{16}} \) with a prime order of 256 and a field size of 1536 [18].

The characteristics of ABE scheme seem to be interesting to be implemented for the user privacy concerns [19] but ABE is a computationally expensive encryption method, especially on resource-constrained devices [20]. We propose, in the next section, a design that gives an end-to-end strong privacy protection with a data-owner centric access control in IoT-cloud environment for e-Health applications.

IV. DESCRIPTION OF THE PROPOSED SOLUTION

The privacy protection is a process that brings several mechanisms at various levels of the data life cycle. To guarantee this protection, it is necessary to use cryptographic primitives for the data encryption and to ensure their confidentiality and integrity. Simultaneously, it becomes mandatory to set up access control mechanisms under data-owner control to enhance privacy. For this purpose, the proposed solution is built on Attribute-Based Encryption and cloudlet architecture (see Fig.3). The proposed solution should be able to provide an access control centred on the data-owner, while preserving privacy in in-home monitoring. The premise of the proposed model provides a total access control for the data owner. Each requester for data must authenticate himself to the data owner while hiding his identity for the cloud. Another advantage of the proposed solution is to reduce the cloud provider intervention. Before detailing our proposed solution, a brief description of the system will be presented. The components of our system are given as roles and the interactions between the roles are presented in Fig.2.

![Fig. 1. CP-ABE working flow](image-url)
TABLE II
SECURITY LEVEL COMPARISON

<table>
<thead>
<tr>
<th>Security Strength (bits)</th>
<th>Symmetric algorithms</th>
<th>RSA-Key length</th>
<th>ABE p (prime order), q (field size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>2TDEA</td>
<td>1024</td>
<td>p=160, q=512</td>
</tr>
<tr>
<td>112</td>
<td>3TDEA</td>
<td>2048</td>
<td>p=224, q=1024</td>
</tr>
<tr>
<td>128</td>
<td>AES-128</td>
<td>3072</td>
<td>p=256, q=1536</td>
</tr>
</tbody>
</table>

Fig. 2. Personal IoT data sharing (For convenience, we omit the first process, which is the initialization)

A. The identified roles in our system

In our system, we identify the following roles: 1) **Data-owner (DO)** is in charge of generating data with his owned devices and storing them in the cloud. The data-owner is the only one who has the right to grant access to his data. 2) **Data-proxy (PR)** in our proposed architecture, the cloudlet acts as a proxy. The cloudlet can be viewed as a base station, a smart box or a WIFI Hotspot/router that hosts a hypervisor. Using a cloudlet instead of a cloud to outsource the computation of the devices fosters the usage of a good bandwidth quality to ensure data security. 3) **Data-requester (RQ)** can be a physician or any other medical practitioner that requests access to the personal data that are well identified by the data owner. To prove the identity of the data requester, we use a PKI infrastructure with an Endorsement Authority. 4) **Endorsement Authority (EA)**: is the trusted authority that verifies the identity and attributes of the requester. The requester provides an evidence of his identity endorsed by this authority. For example, the French digital health agency (ASIP Santé) maintains a directory of the health professionals. 5) **Storage (ST)**: is instantiated by a cloud storage. This actor can only check, if an anonymous requester can provide evidence, which is allowed by DO to access data. 6) **Data-sources (DS)**: are the devices that generate data (sensors or any health devices used to make measurements).

B. Assumptions

In the proposed solution, we make the following assumptions: 1) The Data-owner knows the identity of the Data-requester, which means that the Data-owner (DO) knows the Data-requester (RQ) public key PKRQ. This assumption involves that DO can authenticate signed messages of that Data-requester. 2) The identity of RQ is authenticated with a current available PKI infrastructure. In this case, DO trusts an

Endorsement authority EA, which issues a digital certificate or maintains a list of valid public keys. The same mechanism is available to authenticate storage (ST). 3) Sensors and cloudlet are under physical control of the Data-owner. The delegation of calculations to the cloudlet will not compromise the solution, the related cloudlet security and trust are out of scope of this work. 4) The perfect cryptography and the messages are exchanged over a network that is controlled by Dolev-Yao intruder [6].

C. Proposed protocol

We use in the proposed solution a CP-ABE including four algorithms (Setup, KeyGen, Encrypt, Decrypt) as defined before and a cryptographic hash function H. To enhance the protection of privacy, the identity of the data requester should be hidden. In order to achieve this, every requester generates a public/secret key pair that will be used to perform the access grant request. After endorsement of the requester identity by EA, the data owner issues an anonymous credential for the requester to let him accessing to the ciphertext in the cloud. An overview of the interaction between actors after initialization steps is given in Fig.2 and detailed message exchanges are shown in Fig.4. Table III presents the notation used in the protocol.
We adopt an offline key distribution solution that can be executed by Data-source and Data-proxy. These operations are transmitted to Data-proxy (cloudlet). The result is a master secret key $Msk$ and a public key $P_k$. The public key $P_k$ is as well as the data access policy $P$ are sent to the cloudlet for encryption.

We provision all the connected devices with cryptographic material to secure communication between devices and the cloudlet. This operation is called bootstrapping in IoT [21]. We adopt an offline key distribution solution that can be easily deployed in the proposed use case. In this kind of solution, a symmetric session key $Sk$ is generated after few data exchanges between the data sources and the data proxy [21]. Authors of [10] have described such a solution.

2) **Cloud storage process**: In these steps, all the data generated by Data-source are transmitted to Data-proxy (cloudlet) with metadata through a secure channel. Data-proxy executes $Encrypt(P_k, P, Data)$ and generates a cipher text $C$. $C$ is stored in the cloud in the form of a key-value, where $key = H(C)$ and $value = C$. Data-proxy sends then the key and metadata to Data-owner. Data-owner maintains a table of $(Metadata, Key, P)$, where $Key$ is used to query $C$ in storage $ST$ and $P$ is an access policy.

Message 1: after the establishment of a key session $K_s$, Data-source can send the collected data to Data-proxy. These data are linked with a metadata $(date, time, type of data, etc.)$. $DS → PR: \{Data_{i}, Metadata_{i}\}K_s$

Message 2: in parallel with Message 2, $PR$ executes $Encryption(P_k, P, Data)$ and computes the hash value of the ciphertext to generate $Key$. $C_i = Encrypt(P_k, P, Data_i)$ and $Key_i = H(C_i)$, the message 2 is $PR → ST: \{Key_i, C_i, K_{DO}\}$

Message 3: $PR$ informs Data-owner that new data have been generated by sending to him Metadata encrypted with session key $K_s$ and $Key$ to localize those data in the cloud: $PR → DO: \{Key_i, Metadata_i\}K_s$

3) **Access grant process**: In this process, DO provides a privilege to $RQ$ to access the ciphertext in the cloud (ST) and issues to $RQ$ an appropriate ABE decryption key ($SK_{RQ}$).

We define some data structures to perform this process:

Request query ($RQuery$): is a data structure that $RQ$ can send to $DO$ to formulate the following request:

$$RQuery = \{A_{RQ}, Query, T, RK_{RQ}, \{N\}K_{DO}\}K'_{RQ}$$

Where $A_{RQ} = \{a_0, a_1, .. a_n\} ∈ A$ is a set of $RQ$ attributes, $Query$ is a description of the data requested, $T$ is a grant time validation, $RK_{RQ}$ is a public key generated to be used in this request and $\{N\}K_{DO}$ is a nonce generated by $RQ$ and cyphered with the public key of $DO$.

$GToken$: This data structure is issued by the data owner $DO$ and sent to the $RQ$ to provide the evidence for the data storage (ST) that this requester has a decryption key and has been granted by the data owner to access to the cyphertext $C$ identified by $Key$. This authorization is valid for $T$ time.

$$GToken = \{\{Key, T, h\}K_{ST}, RK_{RQ}, K_{DO}\}K'_{DO}$$

With $h = H(SK_{RQ}||N_{RQ})$. We also define two primitives that will be used:

Endorse $RQ(RQuery)$: This primitive allows $DO$ to parse $RQuery$, checks and authenticates the identity of $RQ$ and his attributes thanks to the Endorsement Authority.

Token $Gen(RQuery)$: Once all the controls are done, $DO$ executes this primitive to generate a structure. This structure is the evidence of the $DO$ agreement to access to the data.

Message 1': $RQ$ sends his query: $RQ → DO: RQuery$. After receiving Message 1', $DO$, executes $Endorse_RQ(RQuery)$ to verify and to authenticate the identity of $RQ$ and his attributes.

Message 2': Once all the controls are done, $DO$ executes $KeyGen(A_{RQ}, Msk)$ algorithm to generate a specific decryption key $SK_{RQ}$ and $Token Gen(RQuery, SK_{RQ})$ primitive to generate an appropriate $GToken$ and to send it to $RQ$: $DO → RQ: \{GToken, \{SK_{QR}\}K'_{RQ}\}$.

4) **Data access process**: When the data owner ($DO$) agrees to grant access to the requester ($RQ$) by sending $GToken$ and $SK_{QR}$, the $RQ$ can ask the cloud (ST) to transfer the ciphered data $C$ in order to decrypt them locally with the received $SK_{RQ}$. $ST$ checks whether $GToken$ is a valid one by performing a $Check_Token(GToken)$ primitive.

Check $Token(GToken)$: $ST$ parses $GToken = \{\{Key, T, h\}K_{ST}, RK_{RQ}, K_{DO}\}$ $DO$ and does the verifications, as presented in table IV.

Message 1'': Once $GToken$ and $SK_{RQ}$ are received, $RQ$ computes $h' = H(SK_{RQ}||N_{RQ})$ and signs it with $RK'_{RQ}$ and sends it to $ST$: $RQ → ST: GToken, \{h'\}RK'_{RQ}$

Message 2'': When $ST$ receives Message 1'', it performs $Check_Token(GToken)$. If all the tests are valid, $ST$ has the evidence that the $GToken$ refers to the data owned by the issuer of $GToken$ and the requester is the one for whom $GToken$ has been issued. At this point, $ST$ can send the cyphertext requested. $ST → RQ: \{C_i\}K'_{ST}$. After Message 2'', $RQ$ can execute $Decrypt$ algorithm to retrieve the clear desired data. The overall workflow of the protocol is given in Fig.4.

V. **MODEL VALIDATION**

In this section, we describe a formal verification of the proposed protocol and we perform an experimental analysis to validate our technical choices.
A. Formal validation

To validate the safety of the proposed system, we made a formal description with a specification of the expected security properties with the High Level Protocol Specification Language (HLPSL) by using the automatic verification tool Avispa [22]. The High Level Protocol Specification Language (HLPSL) is based on roles description. For each role, we define a set of variables to describe the state and transition rules that describe the behavior of the agent who plays this role. There are specific roles: session role, which is a combination of different roles and environment role to define the intruder initial knowledge, the expected security specification in goals section and a session composition execution, that are useful to simulate different attack scenarios. Avispa is a prefferment tool to verify security properties. It is composed from automatic analysis techniques called backends: On-the-fly model-checker (OFMC), CL-based Attack searcher (CL-AtSe), SAT-based model-checker (SATMC), and Tree automata-based protocol analyzer (TA4SP). Avispa input is an HLPSL specification that is converted in a lower level language called Intermediate Format (IF) that can be interpreted by different backends. The system is modeled using HLPSL specification. The behavior of each entity involved in the protocol (DO, PR, DS...) is described as roles. With regard to the e-Health security goals, we define in our model the secrecy of data and metadata and authentication of RQ by DO as security properties. The authentication is implicitly expressed in the protocol (PKI infrastructure) but we explicitly specify an authentication exchanges derived from TLS certificate between the DO and DR with HLPSL. The cloud Gtoken control is specified using a matching variable to define which messages are acceptable. Secrecy of Data-Request identity is modeled with predicate secret(RQ', sec_2, DO, RQ), which means that this identity should be known only by the Data-Owner and the Data-Requester. In the same way, we model secrecy of Data and Metadata. Authentication of Data-Requester by Data-Owner is specified by the predicates request(DO, RQ, do_rq2, Ndo',Nrq) and witness(RQ, DO, do_rq2, Ndo',Nrq), where Ndo',Nrq is the message data on which RQ and DO authenticate. Besides the scenarios explored automatically by the tool, we feigned other possible attacks, where the intruder may play several roles. The results obtained by AVISPA show that the protocol is secure and satisfies the expected security properties, as presented in Fig.5, under assumptions made in IV-B.

B. Experimental analysis

To validate our proposal, we perform experiments with C code on Raspberry Pi platforms acting as constrained devices and on a workstation as a cloudlet node. The experimental simulation of the CP-ABE scheme is done using the pairing based cryptography library (PBC-Library) [23]. We also perform some tests with AES 129 ECB provided by [24] with Raspberry Pi platforms to compare its execution time with the original CP-ABE scheme [25] and with another recent CP-ABE proposition [13]. The workstation runs 64-bits Ubuntu 16.04 LTS, with Intel(R) Core (TM) i5-4590s 3.00GHz CPU and 8GB RAM. The Raspberry Pi 3 Model B runs a Raspbian operating system, with 1.2GHz 64-bit quad-core ARMv8 and 1GB RAM. To achieve a 128-bit security level, we slightly modify the original PBC-Library Type-A pairing parameters to use a 256-bit elliptic curve group based on the supersingular curve $y^2 = x^3 + x$ over 1536-bit finite field. The number of attributes is $N = \{5, 10, 20, 30, 40\}$. We consider this range to be representative of the real-world applications. To avoid errors, the experimental results are the means of 10 trials. For the in-home monitoring use case, we analyze the encryption time because it is the most significant operation performed by the data-owner constrained devices. We simulate the encryption algorithm of CP-ABE scheme [25] both on constrained and unconstrained devices. We take the significant computational operations by considering the number of exponentiation in $G_0$ and $G_T$ and pairing operation. Fig.6 (a) shows the execution time of CP-ABE encryption algorithm on Raspberry Pi platform and workstation. As we can expect, the execution time on Raspberry Pi is in average 12.5 time slower than in a workstation where the time does not exceed one second for up to fourteen attributes. Furthermore, even if the speed of AES algorithm in Raspberry Pi is 18.9 times slower than in the workstation, as we can see in Fig.6 (b), the results presented in table V show that AES is significantly more efficient than ABE in constrained devices. The results motivate our system architecture to outsource ABE encryption to the cloudlet.

VI. Conclusion

In this paper, we presented a data-owner centric approach that provides a limited grant for the cloud provider and that takes into consideration the Internet of things constraints. Our solution enhances privacy protection by masking the data client identity for the cloud service provider. In addition, we...
have performed a formal analysis of the proposed protocol and made an experimental test analysis that demonstrated the efficiency of our cloudlet-based architecture. To enhance security in the IoT side, a complementary work has been done in [26] to propose a computationally intelligent model to measure possible vulnerabilities based on bio-inspired intelligence of ant colony.

### REFERENCES


### TABLE V

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>892.9518</td>
<td>150.50145</td>
</tr>
<tr>
<td>10</td>
<td>1902.03855</td>
<td>110.6723</td>
</tr>
<tr>
<td>20</td>
<td>3310.6896</td>
<td>132.8921</td>
</tr>
<tr>
<td>30</td>
<td>5233.1485</td>
<td>147.55785</td>
</tr>
<tr>
<td>40</td>
<td>6854.60375</td>
<td>153.5809</td>
</tr>
</tbody>
</table>

Fig. 6. Experimental results

(a) CP-ABE encryption time

(b) AES 128 execution time