12-2016

Identity-based Schemes for a Secured Big Data and Cloud ICT Framework in Smart Grid System

Feng Ye
University of Dayton, fye001@udayton.edu

Yi Qian
University of Nebraska-Lincoln

Rose Qingyang Hu
Utah State University

SECURITY AND COMMUNICATION NETWORKS
Security Comm. Networks 2016; **9**:5262–5277
Published online 8 November 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/sec.1695

RESEARCH ARTICLE

Identity-based schemes for a secured big data and cloud ICT framework in smart grid system

Feng Ye[1], Yi Qian[2]* and Rose Qingyang Hu[3]

[1] Department of Electrical and Computer Engineering, University of Dayton, Dayton, OH, U.S.A.
[2] Department of Electrical and Computer Engineering, University of Nebraska-Lincoln, Lincoln, NE, U.S.A.
[3] Department of Electrical and Computer Engineering, Utah State University, Logan, UT, U.S.A.

ABSTRACT

Smart grid is an intelligent cyber physical system (CPS). The CPS generates a massive amount of data for efficient grid operation. In this paper, a big data-driven, cloud-based information and communication technology (ICT) framework for smart grid CPS is proposed. The proposed ICT framework deploys hybrid cloud servers to enhance scalability and reliability of smart grid communication infrastructure. Because the data in the ICT framework contains much privacy of customers and important data for automated controlling, the security of data transmission must be ensured. In order to secure the communications over the Internet in the system, identity-based schemes are proposed especially because of their advantage in key management. Specifically, an identity-based signcryption (IBSC) scheme is proposed to provide confidentiality, non-repudiation, and data integrity. For practical purposes, an identity-based signature scheme is relaxed from the proposed IBSC to provide non-repudiation only. Moreover, identity-based schemes are also proposed to achieve signature delegation within the ICT framework. Security of the proposed IBSC scheme is rigorously analyzed in this work. Efficiency of the proposed IBSC scheme is demonstrated with an implementation using modified Weil pairing over an elliptic curve. Copyright © 2016 John Wiley & Sons, Ltd.

KEYWORDS
smart grid; cyber security; cyber physical system; big data; cloud computing

***Correspondence**
Yi Qian, Department of Electrical and Computer Engineering, University of Nebraska-Lincoln, Lincoln, NE, U.S.A.
E-mail: yi.qian@unl.edu

1. INTRODUCTION

Smart grid is a major evolution of the power grid. Modern control with information and communication technology (ICT) is vastly applied in smart grid. With the advanced ICT infrastructure, two-way communications between customers and service providers is achieved [1–4]. Therefore, smart grid is an advanced cyber physical system (CPS). The CPS generates various types of data by smart meters and different sensors. Because of the massive scale of deployment and real-time (or near real-time) monitoring/controlling requirements, communications in smart grid carry a huge amount of data.

In this paper, we propose an ICT framework for smart grid to cope with big data generated in the CPS. By processing the big data (especially the metering data), smart grid is to achieve efficient and effective demand-response (DR) system [5–7]. DR system is one of the most important components in smart grid. It smooths the power load of the grid so that waste can be reduced in those fossil fuel-based power stations while renewable power sources can be integrated more efficiently. To achieve DR, energy forecast is needed for utility companies to plan power generation and price forecast is needed for customers to manage power usage. In order to achieve high precision, large amount of data is needed and refreshed frequently. Obviously, current control centers provided by utility companies have limited computing capabilities that cannot fulfill the tasks in general. Therefore, in the proposed ICT framework, big data analytics and cloud computing are introduced to assist the utility company. Because of the large scale of smart grid, hybrid cloud service is proposed in the ICT framework. Private local cloud servers are established and maintained by utility companies in a distributed fashion. Smart grid operations rely on the massive amount of data and information are exchanged within the ICT framework.

Data collection and information sharing in smart grid has many security concerns. For instance, metering data

from the customers may contain privacy of individuals [8–10]. Inaccurate control data from service providers may cause extra loss in the power grid, or even a blackout. As a result, metering data is usually transmitted through private networks in smart grid. Monitoring networks for power transmission lines are also private networks [11]. In this case, utility companies shall have more control over the collected data and communication networks. Actions may be taken quickly if abnormal situations occur. However, with the introduction of cloud computing, some data must be transmitted through public networks (e.g., the Internet) [12–14]. Although public cloud service providers have certain security mechanisms within the cloud, data exchange over the Internet still needs extra protection, because a utility company would still mandate to have entire control of the networks within its service area. It is inefficient if not impossible to implement traditional cryptographic systems, symmetric or public, while providing full security control over the data transmission for the utility companies. Identity-based (ID-based) security schemes [15–17] fit here mainly for three reasons: (i) ID-based security schemes utilize the identities of participants to generate public keys so that public key distribution is simplified; (ii) keys and other secret parameters can be updated easily to accommodate status change of participants; and (iii) an authentication center is required to control domain parameters as well as domain secrets. By adopting ID-based security schemes, a utility company shall be in charge of the network security efficiently even if the transmissions are carried over the Internet, which is the case in the proposed ICT framework. Therefore, an ID-based security scheme is proposed in this work to enhance the security of the ICT framework.

The core of the proposed security solution is an ID-based signcryption (IBSC) scheme, which performs the functions of both digital signature and encryption simultaneously. The proposed IBSC scheme utilizes public key cryptography where the public key is computed mainly based on the ID of each participant together with an expiration indicator *time*. As a result, public keys and related domain secrets can be refreshed easily after each session. Furthermore, public keys can be computed locally by any legitimate user in the domain. Therefore, key management is simplified to fit the ICT framework. The proposed IBSC scheme performs efficiently with carefully chosen bilinear pairing operation and other system parameters. Despite its simplicity, the proposed IBSC scheme provides confidentiality, data integrity, and non-repudiation. The IBSC scheme can be reduced to an ID-based digital signature for those cases that do not require confidentiality. In order to enhance the performance, the proposed IBSC is also modified for session key distribution instead of direct message encryption. In addition, the identity-based schemes are also applied to achieve signing right delegation. With this feature, a control center is able to hand its data control to another (or a few other) control center temporarily in the situations of routine maintenance, system failure, and others.

In summary, the main contributions in this work are as follows:

- A big data-driven and cloud-based ICT framework is proposed for smart grid CPS.
- An identity-based signcryption scheme is proposed to secure data transmissions in the proposed ICT framework.
- Signing right delegation from one control center to another (or a few) control center is achieved by identity-based schemes.
- Rigorous security analysis is presented for the proposed IBSC scheme.
- Performance of the proposed IBSC scheme is evaluated with numerical results.

The rest of the paper is organized as follows. In Section 2, related work is discussed. In Section 3, the proposed ICT framework is presented. In Section 4, the proposed identity-based security scheme is illustrated. In Section 5, security of the proposed schemes is analyzed. In Section 6, performance evaluation of the proposed scheme is presented. In Section 7, conclusion and future work are presented.

2. RELATED WORK

Security in smart grid CPS has been widely studied, especially in the area of private networks [1,3,18]. Many of the existing work focused on the advanced metering infrastructure (AMI) because of its importance to DR in smart grid. Metering data collected in the AMI is undoubtedly large in volume and refreshes frequently [1]. With more deployment of renewable energy sources, a large variety of data will be introduced to smart grid additionally. For example, ambient environmental status, energy storage unit status, and weather forecast. Therefore, big data analytics is expected to take action in smart grid [19,20]. Cloud computing has been introduced to smart grid so that big data analytics can take place [10,14,19]. Compared with the frameworks proposed in [10,19], the ICT framework proposed in this work is more comprehensive. There are private networks set by a utility company, a hybrid cloud-based control center with sensitive data collected and pre-processed at local control centers (LCCs) and a visionary idea of harvesting data from various public sources. In addition, this work focuses on providing secure communications for the ICT framework.

The authors in [10] studied a similar framework with security focus. Our work is distinguished in quite a few aspects. First, our proposed security scheme provides encryption and digital signature simultaneously that make it a simpler solution. Second, our proposed scheme has a mechanism to automatically refresh domain secrets (i.e., public keys and other domain public parameters) so that secrets can be easily revoked when a participant leaves the system. Third, more applications such as signing delegation is considered in this work.

Identity-based cryptographic schemes have been widely studied [10,15–17] recently. Unlike well-known symmetric cryptographic schemes (e.g., advanced encryption scheme), ID-based cryptographic schemes are relatively inefficient. Thus, they need to be redesigned or modified for different applications in the proposed ICT framework based on specific requirements. For instance, some data in our framework requires both confidentiality and non-repudiation while the computation needs to be efficient, some data requires non-repudiation only, the domain secrets need to be refreshed frequently, and others.

3. INFORMATION AND COMMUNICATION TECHNOLOGY FRAMEWORK FOR SMART GRID

3.1. Overview of the information and communication technology framework

Figure 1 depicts an overview of the proposed ICT framework. Three types of networks are applied in this framework, including local area networks (LANs) established by customers, private networks established by utility companies (or service providers), and the Internet provided by a third-party Internet service provider. The combination of the aforementioned networks establish two-way communications between utility companies and customers. There are four parties in the framework, namely, internal data collectors (i.e., customers and grid monitoring sensors), a service provider, power generators, and external information sources. The first three parties are directly related to smart grid. External information sources do not belong to smart grid, nonetheless they provide insightful information to smart grid operations.

3.2. Networks in the information and communication technology framework

Communications in the ICT framework are achieved through both private networks deployed by utility companies and the Internet. Specifically, internal data is gathered and transmitted to utility companies through private networks. For example, metering data is uploaded through the AMI. The wide area monitoring system for phasor measurement unit (PMU) data consists of private networks. Monitoring networks for power transmission lines are also private networks [11].

Security is one important reason for deploying private networks. For instance, metering data is gathered from customers; thus, it contains much private and sensitive information of customers. Life style of a customer may be revealed if metering data is leaked. In the AMI, a home area network is established within a household, connecting a smart meter and smart appliances with sensors and actuators. Each smart meter uploads data to a data aggregate unit. data aggregate units in a neighborhood form a neighborhood area network. Metering data finally reaches the metering data management system through a high speed backhaul wide area network.

Reliability is another important issue. For instance, real-time monitoring in smart grid requires low latency (e.g.,

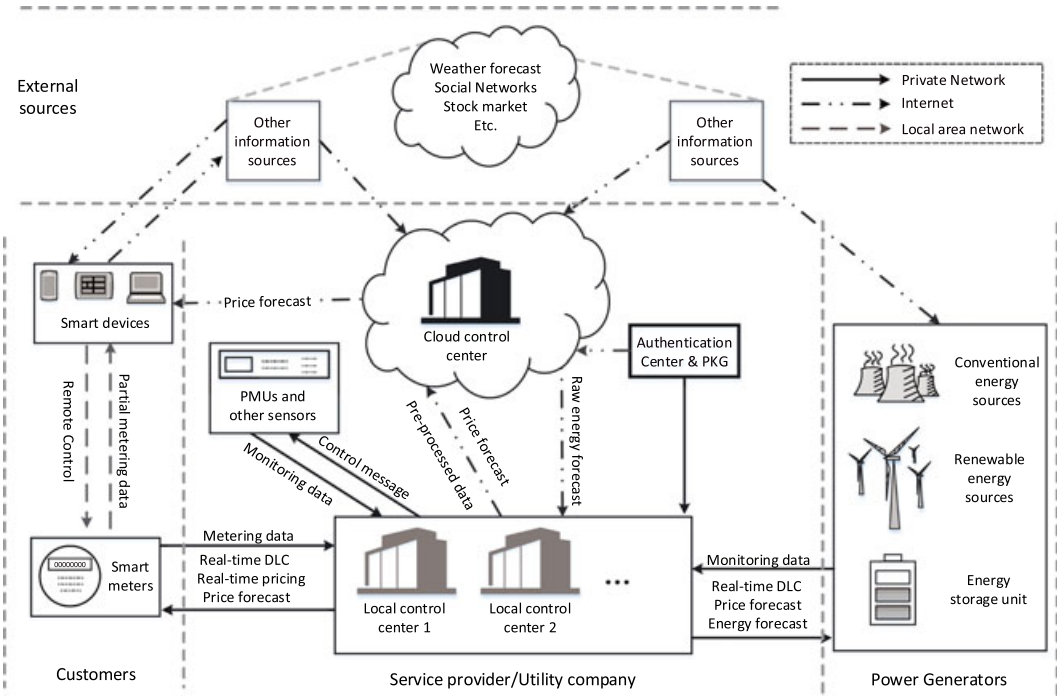

Figure 1. An overview of the proposed information and communication technology (ICT) framework.

DOI: 10.1002/sec

10 to 100 ms for PMUs in wide area monitoring system). Public cellular networks or the Internet service providers can hardly achieve the latency requirement because of their complicated protocols and mechanisms. The private networks in smart grid are designed specifically to fulfill the latency requirement.

Cost is also a major reason for private networks. Subscription fees for public network service providers can be overwhelmingly high in the long run. Moreover, power grid may cover larger areas than what a public network service covers. For instance, power transmission line monitoring networks may cover remote areas that have no public network access. The private networks are built on different types of communication technologies, including various types of wireless networks (e.g., Wi-Fi, Zigbee, and WiMAX) and high speed wired networks (e.g., optical fiber and Ethernet). Specific technologies are chosen to balance optimal network performance with hardware/maintenance cost.

3.3. Internal data collectors

Internal data is collected by sensors deployed in the power grid. Specifically, smart meters are deployed at the customer side by utility companies. Many types of sensors are deployed by utility companies to monitor transmission lines, substations, and others. In smart grid, customers are motivated to actively participate in DR. Therefore, some of the information is given to customers while being uploaded to utility companies. For instance, customers are granted access to electricity usage data of their own properties. In many cases, customers have self-established LANs or wireless LANs (WLANs), for example, Wi-Fi and bluetooth, which connect smart devices, smart appliances, and corresponding smart meters in each household. Smart devices can be smart phones, tablets, laptops, and so on. Without loss of generality, smart phones will be used for the indication of smart devices. From smart phones, customers are able to monitor electricity consumption of their appliances. With network access, customers have remote control capability over their appliances. In many cases, LANs/WLANs established by customers have access to the Internet. Consequently, remote monitoring and remote control can be applied by customers anywhere with Internet access. Apart from fixed household appliances, electrical vehicles (EVs) and plug-in hybrid EVs are mobile appliances that have more resilient electricity requirements. For example, an EV can be charged in a household, it can also be charged in a public charging station or a capable parking lot. Furthermore, some companies and researchers are pushing to standardize batteries for EVs. In that case, customers can go to battery exchange station and replace an empty battery with a fully charged one. The electricity consumption of EVs at charging ports will be captured by smart meters. Other useful information such as location and possible routes of EVs may be gathered by some external agents with permission of the customers. Such data will be considered as external data to the ICT framework.

3.4. Control centers and power generators

Control centers are deployed and operated by utility companies. Specifically, there are three types of control centers in the proposed framework, LCCs, a cloud control center (CCC), and an authentication server (AS) with a private key generator (PKG). For simplicity, AS or PKG will be used interchangeably indicating the group hereafter. It is also reasonable to assume that PKG is a trusted third party. In terms of responsibility, control centers as a whole unit makes *energy forecast* to power generators and makes *price forecast* to customers. In demand response, a control center is also responsible of direct load control over both power consumption from customers and electricity generation from power generators. Direct load control usually apply to power companies themselves. For instance, they cycle air conditioners(ACs) and water heaters on and off during periods of peak demand to smooth the power generation.

Undoubtedly, the power grid has a large scale. Therefore, LCCs need to be distributed across the power grid for better scalability and reliability in the proposed ICT framework. For instance, an LCC can be deployed close to or inside a power distribution substation. While each substation covers a relatively small area, an LCC is responsible for the customers within that particular area. Substations are currently connected by a high-speed private backhaul network deployed by the utility company. With extra gateways to the backhaul network, LCCs are interconnected reliably. Although LCCs and private networks are controlled by utility companies, it is safer to assume that LCCs do not share collected data during normal operations. Some necessary conditions for a LCC to share or pass data to other LCCs include routine maintenance, temporary system off-line because of cyber attacks or natural disasters. Main functions of LCCs include (i) internal data (i.e., metering data and sensor data) collection, (ii) preprocessing for sensitive data so that privacy of customers is protected, (iii) real-time direct control of the power grid when needed, (iv) finalizing energy forecast for power generators, and (v) generating price forecast for customers.

Different from LCCs, a CCC is a comprehensive unit comprised with complicated and distributed hardware as well as software. Different levels of services such as infrastructure as a service, platform as a service, and software as a service can be provided. Nonetheless, such complexity needs to be transparent to customers (i.e., the utility company). Therefore, the CCC is viewed as a powerful single unit in the framework. The CCC is provided by a public cloud service provider. It is connected internally by high speed networks that are not controlled by the utility company. For instance, the Internet and private networks of cloud service providers. Main functions of CCC include (i) store data uploaded from the LCCs for a certain period, (ii) fetch data from external sources, and (iii) perform big data analytics to collected data and make raw energy forecast for each area and the entire grid. Raw energy forecast is sent back to LCCs for finalizing.

Power generators consist of conventional energy sources and renewable energy sources. On one hand, if energy forecast is provided for a sufficient time period,

conventional power generators are able to optimally control fuel consumption. Transition between peak and off-peak electricity generations can be more efficiently. On the other hand, renewable energy sources appear to be less profitable from energy forecast because of uncontrollable sources. However, some kind of renewable resources are predictable. For example, solar power is predictable with accurate weather forecast. Therefore, in order to control power generation optimally, estimated capacity of electricity generation from renewable resources is fed back to the service provider. A better energy forecast for conventional power generators can be made after receiving that information. Energy forecast in different granularity shall be updated based on the latest collected data and the results from data analysis.

3.5. External data sources

From the discussion earlier, we can see that, internal data alone is not enough to make accurate energy forecast for conventional power generators and energy storage units. Different types of data from external sources are included in the ICT framework. As mentioned earlier, for instance, weather forecast can provide better estimation of electricity generation from renewable energy sources. Locations and routes of EVs can be used to estimate energy consumption as well as the schedule. Useful information can also be extracted in many other external sources. Smart grid will certainly operate more efficiently with all those external data.

3.6. Big data and cloud computing in smart grid

Forecasting function of the ICT in smart grid is depicted in Figure 2. It includes data input, big data analytics, and information output. Input data of the ICT infrastructure consists of internal data generated from smart grid and external data from other sources. Internal data is generated from the infrastructures in smart grid. For example, metering data is generated frequently (e.g., every 15 min) by smart meters deployed in AMI. Metering data reveals the electricity usage of the power grid. It is important for the service provider to adjust electricity generations from the power generators. Besides metering data, the monitoring and control system in smart grid also generates various monitoring data from different sensors, such as PMUs and transmission line monitoring sensors. Sensing data that reveals the operational status of the power grid is generated in real time (e.g., PMU generates data at high frequency, for example, 60 to 120 frames per second for 60-Hz system [21]). It is important for the service provider to be aware of any abnormal situation in the power grid in real time. Therefore, actions can be taken to prevent a blackout or to quickly recover from a blackout. With internal data, the service provider is able to monitor the power grid and take actions if necessary.

External data from other sources is also an important input for the ICT infrastructure. For example, the electricity to be generated from conventional power generators depends not only on energy requirements from customers but also from the capacity of renewable sources and storage units. The capacity of renewable sources (e.g., a solar farm) is not likely controllable. Nonetheless a precise weather forecast will be helpful for the predication of that capacity. There are various types and sources of external data. For example, it can be weather forecast, data from social networks, data from location-based tracking applications of smart devices, and many others.

The collected data assists the service provider to optimize the control over the power grid, such as giving energy forecast in different granularity (e.g., daily, hourly, and per minute) to power generators and price forecast in different granularity to customers and smart appliances. In order to achieve the optimal control, the service provider needs to perform big data analytics in four steps, data collection, data pre-processing, data storage, and data analysis [22].

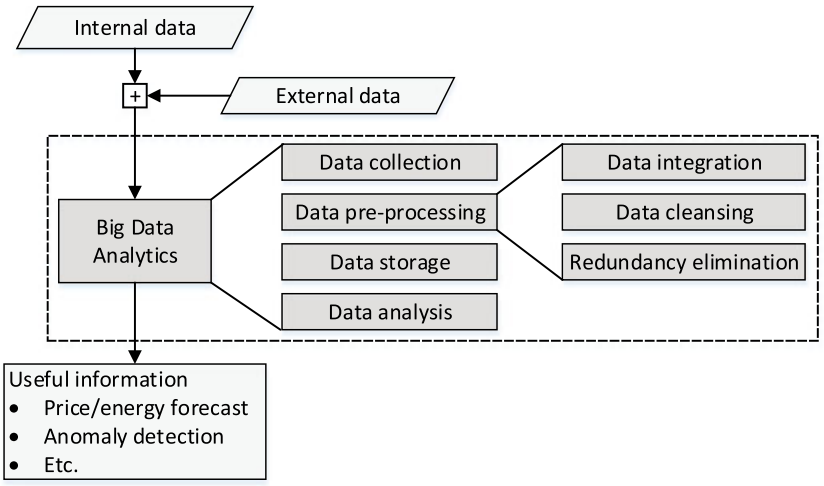

Figure 2. Data processing procedure.

DOI: 10.1002/sec

For instance, the service provider models the energy consumption of customers, and more importantly, the schedule of their appliances. The energy consumption of each appliance may depend on its load, for example, an air conditioner uses more electricity if lower temperature is set. However, the fluctuation of power is not too much for many appliances (e.g., washing machine and coffee machine). Besides, the schedule and energy consumption is associated to surrounding environments, for example, temperature and humidity. Furthermore, big data analytics in social networks is being widely studied [23] as well. Useful information can be mined from the status posted by customers to assist smart grid operations.

However, enormous computing and storage resources are needed to extract useful information efficiently. How can utility companies achieve it with a reasonable budget? A feasible solution is to introduce cloud computing into the ICT infrastructure [24]. Cloud computing brings three major advantages to smart grid. Firstly, there is no need to invest on the whole infrastructure, only the infrastructure for the private portion of the cloud is deployed by utility companies. The resources from the public portion of the cloud can be rent at a relatively low price. In many cases, cloud computing uses a pay-as-you-go pricing model. The maintenance cost is also low because it only applies to the private portion. Secondly, it is easier to implement applications in cloud. Public cloud computing has virtually unlimited resources. Therefore, the utility companies need not worry about upgrading capacity for large-scale system, which is no doubt a huge concern in smart grid. The infrastructure in cloud can also be rescaled according to adaptive requirements. Moreover, because of the elasticity, feature updates/upgrades can be performed in a short amount of time without disturbing users to install major updates or extra packages. Thirdly, it is easier to access cloud service from a variety of smart devices. Because of that, monitoring and controlling of the grid can be more flexible. Once the security is provided over the transmission, cloud service can be accessed from virtually anywhere with authorization.

3.7. Security requirement of the framework

There are different types of information frequently transmitted in the proposed ICT framework. General security requirements are listed in Table I for each type of information.

Because LANs are established for customers, the security is protected by corresponding wireless protocols. Because the utility company have control over the private networks, information exchanged through private networks is likely well protected. However, the utility company have no control over Internet or the CCC; therefore, communications between LCCs and the CCC need more security protection in addition to the default protection provided by Internet protocol and cloud computing service providers. Moreover, public cloud service may not fulfill the security requirements from the utility company. Therefore, we propose to apply identity-based security schemes for power companies to secure the data transmission in the aforementioned framework.

3.8. Applications of the proposed security schemes

Encryption and digital signature: The proposed security schemes can be applied directly to provide both confidentiality and non-repudiation. For instance, pre-processed metering data sent from LCCs to the CCC is encrypted and signed to provide confidentiality and non-repudiation. Information generated by big data analytics is also encrypted and signed before being sent from the CCC to LCCs.

Session key distribution: If symmetric ciphers are preferred in some applications, the proposed identity-based scheme can be applied to achieve secure session key distribution.

Signing right delegation from L_i to L_j: If L_i is subject to a routine maintenance, it may delegate signing rights to another LCC (e.g., L_j). As shown in Figure 3, the PKG has the authority to delegate signing right from L_i to L_j. Alternatively, L_i can delegate signing right to L_j locally without involving another party for more efficient operation.

Signing right delegation from L_i to a group of *LCC*s: The PKG can assign a group of LCCs as a group proxy to sign for L_i, as illustrated in Figure 4. Without loss of generality, assuming a total number N LCCs are chosen as a group in the rest of the paper.

Table I. Security requirements.

	Confidentiality	Data Integrity	Non-repudiation
Metering data	✓	✓	
Monitoring data		✓	
Control message		✓	✓
Raw energy forecast	✓	✓	✓
Pre-processed data	✓	✓	✓
Price forecast		✓	✓
Energy forecast	✓	✓	✓
Other information		✓	

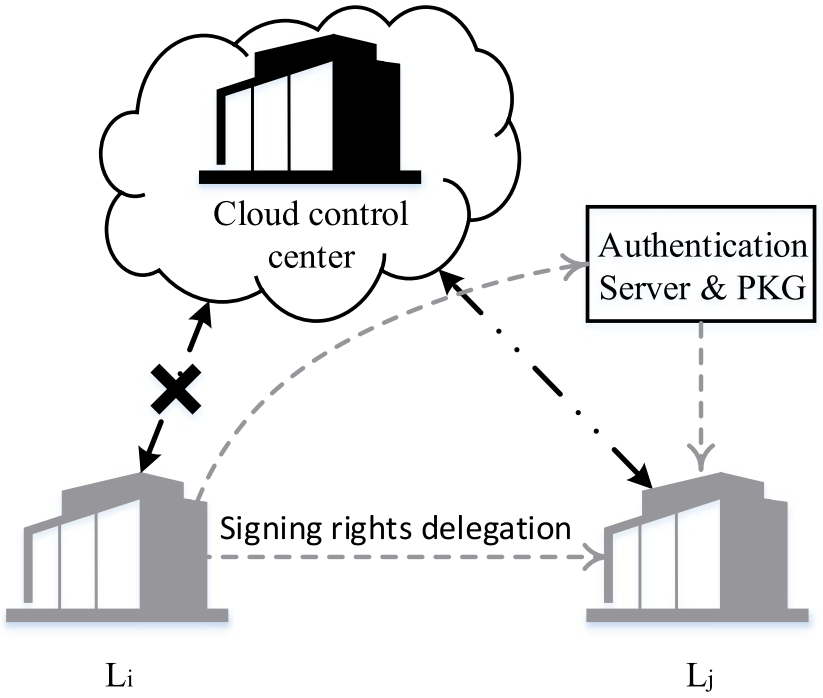

Figure 3. Signing right delegation from L_i to L_j.

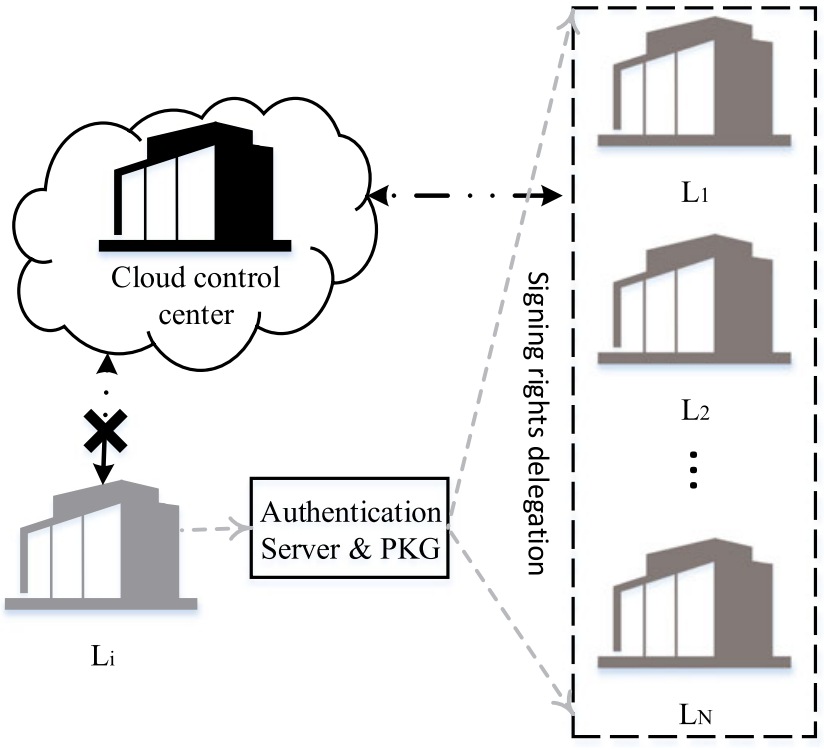

Figure 4. Signing right delegation from L_i to a group of local control centers (LCCs).

4. IDENTITY-BASED SECURITY SCHEMES

In the ICT framework, each component has a unique ID. The foundation of ID-based security scheme is public key cryptography. However, instead of generating keys randomly, ID-based security scheme utilizes the unique ID of each participant. By doing so, key management might be more convenient because some of the keys can be computed locally or even ahead of time. Furthermore, privacy and authentication can still be provided to the participants.

In the proposed ID-based security, we adopt ($ID||time$) instead of ID for public key generation, where $time$ is the expiration time of current session. In the next session, the entire participants will update corresponding secrets and parameters accordingly. When a participant leaves the system domain, secrets bared by this participant need to be revoked. By adopting $ID||time$, if the PKG stops issuing secret keys to the left participant, key revocation can be performed automatically at the beginning of the next session. New messages will not be disclosed to old keys.

ID-based security scheme has several applications in the ICT framework. For example, privacy of the messages (e.g., pre-processed data) sent from LCCs to the CCC can be protected by ID-based encryption. Authentication of the messages sent from control centers to customers (e.g., pricing forecast) can be protected by ID-based digital signature.

4.1. Preliminaries

The proposed ID-based security scheme is based on bilinear map. Let $\mathbb{G}_1$ and $\mathbb{G}_2$ be groups of prime order q. Let g be a generator of $\mathbb{G}_1$. Let $\hat{e}: \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$. We say that $(\mathbb{G}_1, \mathbb{G}_2)$ are bilinear map groups if $\hat{e}$ has the properties in the following:

- Bilinearity: $\hat{e}(aP, bQ) = \hat{e}(P, Q)^{ab}$ for all $P, Q \in \mathbb{G}_1$ and all $a, b \in \mathbb{Z}_q^*$.
- Non-degeneracy: For any $P \in \mathbb{G}_1$, $\hat{e}(P, Q) \neq 1$ for all $Q \in \mathbb{G}_1 \backslash \{\mathcal{O}\}$ (indicated as $\mathbb{G}_1^*$ hereafter).
- Computability: There is a polynomial time algorithm for computing $\hat{e}(P, Q)$ for all $P, Q \in \mathbb{G}_1$.

4.2. Identity-based signcryption

The proposed IBSC scheme comprises five algorithms: *Setup*, *Keygen*, *Signcrypt*, *Decrypt*, and *Verify*. Without loss of generality, let A ($ID_A = A$) sends message $M = \{0,1\}^n$ to B ($ID_B = B$). The detailed IBSC scheme is described in the following:

Setup: The PKG chooses groups $(\mathbb{G}_1, \mathbb{G}_2)$ of prime order q, a generator g of $\mathbb{G}_1$, a randomly chosen master key $s \xleftarrow{R} \mathbb{Z}_q^*$, a domain secret $g_1 = sg \in \mathbb{G}_1$. The PKG also chooses three cryptographic hash functions,

$$H_1: \{0,1\}^* \to \mathbb{G}_1^*,$$
$$H_2: \{0,1\}^* \to \mathbb{Z}_q^*,$$
$$H_3: \{0,1\}^* \to \{0,1\}^n$$

The domain public parameters are

$$params = \langle \mathbb{G}_1, \mathbb{G}_2, g, q, g_1, H_1, H_2, H_3, n \rangle$$

The public/private keys of the AS are $p_{AS} = H_1(AS||time)$ and $d_{AS} = sp_{AS}$.

Keygen: For a given string $ID \in \{0,1\}^*$ and a expiration time stamp $time$, the algorithm builds a public/private key pair p_{ID}/d_{ID} as follows.

- Public key: $p_{ID} = H_1(ID||time)$.
- Private key: $d_{ID} = sp_{ID}$.

Note that $time$ is converted into $\{0,1\}^*$ and is concatenated to ID in the illustration. Other processes can be taken for the same purpose, for example, $time$ can also be XORed to ID.

Signcrypt: To signcrypt a message M, sender A

(1) randomly picks $r \xleftarrow{R} Z_q^*$ and sets

$$U = rg$$

(2) computes $h_1 = H_2(M||A||U)$ and sets

$$V = d_A h_1 + rg_1$$

(3) computes $p_B = H_1(B||time)$ and $h_2 = H_2(A||B)$, and sets

$$X = h_2 U$$

(4) computes $h_3 = H_3\left(X||\hat{e}(rg_1, h_2 p_B)\right)$ and encrypts the message

$$W = M \oplus h_3$$

(5) finally outputs a 4-tuple $\langle U, V, X, W \rangle$.

Note that in the 4-tuple, $\sigma = \langle U, V \rangle$ serves as the digital signature and $C = \langle X, W \rangle$ serves as the cipher text.

Decrypt: Upon receiving $\langle \sigma, C \rangle$, receiver B decrypts M in the steps as follows.

(1) B computes $h_3' = H_3\left(X||\hat{e}(X, d_B)\right)$;
(2) decrypts $M = W \oplus h_3'$.

Verify: With M recovered, B continues to verify the digital signature in the steps as follows.

(1) computes $p_A = H_1(A||time)$, and $h_1 = H_2(M||A||U)$;
(2) verifies if $\hat{e}(g, V) = \hat{e}(g_1, p_A h_1 + U)$.

This completes the description of the IBSC scheme. We then verify consistency. To *Decrypt* M, it must be $h_3' = h_3$. It can be verified as follows:

$$\begin{aligned} \hat{e}(X, d_B) &= \hat{e}(h_2 rg, sp_B) \\ &= e(rg, p_B)^{sh_2} \\ &= e(rg_1, h_2 p_B) \end{aligned} \tag{1}$$

Therefore,

$$\begin{aligned} h_3' &= H_3\left(X||\hat{e}(X, d_B)\right) \\ &= H_3\left(X||\hat{e}(h_2 rg, sp_B)\right) \\ &= h_3 \end{aligned} \tag{2}$$

Algorithm *Verify* is consistent because

$$\begin{aligned} \hat{e}(g, V) &= \hat{e}(g, d_A h_1 + rg_1) \\ &= \hat{e}(g, p_A h_1 + rg)^s \\ &= \hat{e}(g_1, p_A h_1 + U) \end{aligned} \tag{3}$$

From the illustration, we can see that sender A encrypts the message with p_B so that confidentiality is provided. Sender A also signs the message with d_A so that non-repudiation is provided. Data integrity is also provided with hash functions.

4.3. Identity-based signature

As discussed in previous section, not all messages need encryption. Nonetheless, data integrity and non-repudiation are still required. Therefore, the IBSC scheme may be reduced to an identity-based signature scheme. The IBS scheme comprises of two algorithms, *Sign* and *Verify*. For consistency, let A sends M to B in the illustration as follows:

Sign: For a given message M, sender A

(1) randomly picks $r \overset{R}{\leftarrow} Z_q^*$ and computes

$$U = rg$$

(2) computes $h_1 = H_1(M||A||U) \in \mathbb{Z}_q^*$ and sets

$$V = d_A h_1 + rg_1$$

(3) finally outputs $\sigma = \langle U, V \rangle$.

Verify: At the receiver side, B

(1) computes $p_A = H_1(A||time)$ and $h_1 = H_1(M||A||U)$;
(2) verifies if $\hat{e}(g, V) = \hat{e}(g_1, p_A h_1 + U)$.

This completes the description of the IBS scheme. The consistency is proven by Equation (3).

4.4. Key distribution and symmetrical cryptography

Although encryption is achieved in the IBSC scheme, some may still prefer symmetric ciphers (e.g., advanced encryption standard) for data encryption. Because the proposed identity-based schemes are based on bilinear pairing (over elliptic curves) with large numbers, they are considerably slow compared with well-established symmetric ciphers. Therefore, the IBSC can be modified for session key distribution with symmetric ciphers (e.g., $E_K(\cdot)$) for the actual data encryption.

Modified IBSC: To secure a message M with a session key K, sender A

(1) randomly picks $r \overset{R}{\leftarrow} Z_q^*$ and sets

$$U = rg$$

(2) computes $h_1 = H_2(M||K||A||U)$ and sets

$$V = d_A h_1 + rg_1$$

(3) computes $p_B = H_1(B||time)$ and $h_2 = H_2(A||B)$, and sets

$$X = h_2 U$$

(4) computes $h_3 = H_3\left(X||\hat{e}(rg_1, h_2 p_B)\right)$ and encrypts the message

$$W = K \oplus h_3$$

(5) encrypts M as $C = E_K(M)$;
(6) finally outputs a 5-tuple $\langle U, V, X, W, C \rangle$.

Digital signature is provided in the same way that IBSC does. The consistency of the modified IBSC follows the original scheme.

4.5. Single proxy signing right delegation

Certificate distribution: Let c_{ij} be the certificate of signing right delegated by L_i to L_j. A simple example of such certificate could be $c_{ij} = AS||j||t_{ij}$, where t_{ij} be the expiration time of c_{ij}. A certificate can be valid for one message, or for all messages before expiration of the certificate. To delegate c_{ij} for a message M, the LCC L_i

(1) randomly picks $y \xleftarrow{R} \mathbb{Z}_q^*$ and sets

$$U = yg$$

(2) computes $h = H_1(M \| c_{ij})$ and sets

$$V = hUp_j$$

(3) sets $W = hd_i + yg_1$.

Signing rights delegation is a 4-tuple $\sigma_c = \langle U, V, W, c_{ij} \rangle$.

Once L_j receives the σ_c, it verifies if $\hat{e}(V, W) = \hat{e}(hUd_j, hp_i + U)$. The consistency is shown in the following:

$$\begin{aligned} \hat{e}(V, W) &= \hat{e}(hUp_j, hd_i + yg_1) \\ &= \hat{e}(hUp_j, hp_i + yg)^s \\ &= \hat{e}(hUd_j, hp_i + U) \end{aligned} \tag{4}$$

Single proxy signature: With certificate c_{ij}, L_j is ready to sign message M on behalf of L_i. To do so, L_j

(1) randomly picks $z \xleftarrow{R} \mathbb{Z}_q^*$ and computes

$$\begin{aligned} \mu &= zg, \\ \xi &= H_1(m \| w \| c_{ij}), \\ \omega &= w + \xi d_j + zg_1 \end{aligned}$$

(2) finally outputs a 5-tuple $\sigma_{ij} = \{c_{ij}, u, w, \mu, \omega\}$.

The proxy signature is $\sigma_{ij} = \{c_{ij}, u, w, \mu, \omega\}$ (note that u, c_{ij}, and w are from L_i). A receiver verifies σ_{ij} by checking if

$$\hat{e}(g, \omega) = \hat{e}\left(g_1, hp_i + u + \xi p_j + \mu\right)$$

The consistency is shown in the following:

$$\begin{aligned} \hat{e}(g, \omega) &= \hat{e}(g, w + \xi d_j + zg_1) \\ &= \hat{e}(g, hp_i + yg + \xi p_j + \mu)^s \\ &= \hat{e}\left(g_1, hp_i + u + \xi p_j + \mu\right) \end{aligned} \tag{5}$$

Signing right delegation by the PKG: Alternatively, the PKG is able to distribute a certificate c_{ij} to L_j. To do so, the PKG

(1) randomly picks $y \xleftarrow{R} \mathbb{Z}_q^*$ and computes

$$\begin{aligned} u' &= yg, \\ h' &= H_1(m \| c_{ij}), \\ v' &= hup_j, \\ w' &= hd_i + hd_{AS} + yg_1 \end{aligned}$$

(2) finally outputs a 5-tuple $\sigma_c' = \langle u', v', u', w', c_{ij} \rangle$.

The delegation σ_c' is verified by L_j if $\hat{e}(v', w') = \hat{e}(h'u'd_j, h'p_i + h'p_{AS} + u')$. The consistency is shown in the following:

$$\begin{aligned} \hat{e}(v', w') &= \hat{e}(h'u'p_j, h'd_i + h'd_{AS} + yg_1) \\ &= \hat{e}(h'u'p_j, h'p_i + yg)^s \\ &= \hat{e}(h'u'd_j, h'p_i + h'p_{AS} + u') \end{aligned} \tag{6}$$

4.6. Group proxy signing right delegation

Group proxy signing right of L_i is delegated by the PKG to a chosen group of LCCs (e.g., L_n for some n).

Certificate distribution: For each L_n, the PKG generates a partial signing right certificate c_{in} and

(1) randomly picks $y_n \xleftarrow{R} \mathbb{Z}_q^*$ and computes

$$\begin{aligned} u_n &= y_n g, \\ h_n &= H_1(m \| c_{in}), \\ v_n &= h_n u p_n, \\ w_n &= h_n d_{AS} + y_n g_1 \end{aligned}$$

(2) finally outputs a 5-tuple $\sigma_n = \langle u_n, v_n, w_n, c_{in} \rangle$.

Once L_n receives the σ_n, it verifies the certificate by checking if $\hat{e}(v_n, w_n) = \hat{e}(hud_j, hp_{AS} + u_n)$.

Partial signature: With σ_n, L_n can generate a partial signature for message M. To do so, L_n

(1) randomly picks $z_n \xleftarrow{R} \mathbb{Z}_q^*$ and computes

$$\begin{aligned} \mu_n &= z_n g, \\ \xi_n &= H_1(m \| w_n \| c_{in}), \\ \omega_n &= w_n + \xi_n d_n + z_n g_1 \end{aligned}$$

(2) finally outputs a 5-tuple $\sigma_n' = \langle c_{in}, u_n, w_n, \mu_n, \omega_n \rangle$.

Group signature: After all the proxies have generated partial signatures, one of the LCCs is chosen as the gateway (e.g., L_j). To generate a group signature, L_j

(1) computes

$$\begin{aligned} \mu_g &= \sum_{n=1}^{N} \mu_n, \\ \omega_g &= \sum_{n=1}^{N} \omega_n, \\ w_g &= \sum_{n=1}^{N} w_n \end{aligned}$$

(2) finally outputs $\sigma_g = \langle \mu_g, \omega_g, w_g, w_n \& c_{in} \ \forall n \rangle$.

Upon receiving the group signature σ_g, a receiver verifies by checking if

$$\hat{e}(g, \omega_g) = \hat{e}\left(g_1, \sum_{n=1}^{N} (h_n p_{AS} + \xi_n p_n) + u_g + \mu_g\right)$$

The consistency can be verified such that

$$\begin{aligned}\hat{e}(g, \omega_g) &= \hat{e}\left(g, \sum_{n=1}^{N} (w_n + \xi_n d_n + z_n g_1)\right) \\ &= \hat{e}\left(g, \sum_{n=1}^{N} (h d_{AS} + y_n g_1 + \xi_n d_n + z_n g_1)\right) \\ &= \hat{e}\left(g_1, \sum_{n=1}^{N} (h_n p_{AS} + y_n g + \xi_n p_n + z_n g)\right) \\ &= \hat{e}\left(g_1, \sum_{n=1}^{N} (h_n p_{AS} + \xi_n p_n) + u_g + \mu_g\right)\end{aligned}$$

5. SECURITY ANALYSIS OF THE PROPOSED SCHEMES

5.1. Assumptions for security analysis

The security of the IBSC and IBS schemes is based on the following computational problems [16,17,25]:

Computational Diffie–Hellman (CDH) problem: given P, aP, and $bP \in \mathbb{G}_1$, to compute $abP \in \mathbb{G}_1$ in polynomial time.
Bilinear Diffie–Hellman (BDH) problem: given P, aP, bP, and $cP \in \mathbb{G}_1$, to compute $\hat{e}(P, P)^{abc} \in \mathbb{G}_2$ in polynomial time.

Without loss of generality, time stamp *time* is viewed as part of the identity *ID* in the analysis later. For better illustration, the proposed IBSC scheme is separated into identity-based encryption scheme and identity-based signature for security analysis. Moreover, all the random values are picked uniformly unless specified.

5.2. Identity-based encryption security

Security models

Definition 1 ((Semantic security for IBE schemes) [25]). *if no probabilistic polynomial time adversary has a non-negligible advantage in this game:*

(1) The challenger runs the setup algorithm to generate the system's parameters and sends them to the adversary.
(2) The adversary $\mathcal{A}$ performs a series of queries:

- *Key extraction queries: $\mathcal{A}$ produces an identity ID and receives the private key d_{ID}.*
- *Challenge: After a polynomial number of queries, $\mathcal{A}$ outputs two equal-length plaintexts M_0 and M_1 and a public key ID on which it wishes to be challenged (ID has not appeared in private key queries). The challenger picks a random bit $T \in \{0, 1\}$ and encrypts M_b according to the IBE scheme.*
- *More key extraction queries: $\mathcal{A}$ issues more key extraction queries. The challenger responds as before.*

Finally, the adversary $\mathcal{A}$ outputs a guess $T' \in \{0, 1\}$, $\mathcal{A}$ wins the game if $T = T'$.

Security analysis

Lemma 1. *Let H_1 be a random oracle from $\{0, 1\}^*$ to $\mathbb{G}_1^*$. An adversary $\mathcal{A}$ has ϵ advantage against IBE. Then there exists an adversary $\mathcal{F}$ that has advantage*

$$\epsilon_{\mathcal{F}} \geq \frac{\epsilon}{e(1 + q_E)} \tag{7}$$

Proof. Let T be a random variable following Bernoulli distribution, that is, $T \in \{0, 1\}$ with probability δ of being 0 and probability $1 - \delta$ of being 1.

Setup: the challenger generates system parameters and sends it to $\mathcal{F}$. $\mathcal{F}$ picks a random value $p_{ID} \xleftarrow{R} \mathbb{G}_1^*$. Then $\mathcal{A}$ issues H_1 queries.
Public key queries on H_1: $\mathcal{A}$ queries oracle H_1 at ID_i, $\mathcal{F}$ responds that,

- If $H_1(ID_i)$ exists in L_1 (which is a list kept by $\mathcal{F}$), then returns that stored value.
- If $H_1(ID_i)$ does not exist in L_1, then $\mathcal{F}$ randomly choose $T \in \{0, 1\}$, and $v_i \xleftarrow{R} \mathbb{Z}_q^*$ and sets

$$p_i = H_1(ID_i) = \begin{cases} v_i g, & T = 0 \\ v_i p_{ID}, & T = 1 \end{cases} \tag{8}$$

$\mathcal{F}$ stores $\langle ID_i, p_i, v_i, T \rangle$ in L_1.
Challenge: $\mathcal{A}$ outputs ID_c and two equal-length messages M_0 and M_1. $\mathcal{F}$ gives the challenger M_0, M_1, the challengers randomly picks $u \in \{0, 1\}$ and encrypts $M_u \to C = \langle X, W \rangle$. $\mathcal{F}$ runs the response to H_1 queries to find p_c such that $H_1(ID_c) = p_c$. Then, $\mathcal{F}$ finds $\langle ID_c, p_c, v_i, T \rangle$.

- If $T = 0$, then $\mathcal{F}$ aborts.
- If $T = 1$, then we have $p_c = v_i p_{ID}$. Set $C' = \langle v_i^{-1} X, W \rangle$. Because

$$\begin{aligned}\hat{e}(v_i^{-1}X, d_c) &= \hat{e}(X, v_i^{-1}sp_c)\\ &= \hat{e}(X, v_i^{-1}sv_ip_{ID})\\ &= \hat{e}(X, d_{ID})\end{aligned}$$

therefore, C' is IBE of M_u under ID_c, and decryption of C' using d_c is the same as decryption of C using d_{ID}.
Guess: $\mathcal{F}$ outputs a guess u'.

If $\mathcal{F}$ does not abort during the process, then $|Pr[u = u'] - 1/2| \geq \epsilon$, where the probability is over the random bits used by $\mathcal{A}$, $\mathcal{F}$ and the challenger. Let

- E_1 be the event that $\mathcal{F}$ aborts in private key queries.
- E_2 be the event that $\mathcal{F}$ aborts in challenge stage.

Then we have the probability of not aborting is

$$Pr(\neg E_1 \wedge \neg E_2) = Pr(\neg E_1)Pr(\neg E_2) = \delta^{q_E}(1-\delta) \quad (9)$$

The maximum probability is achieved at $\delta_{opt} = \frac{q_E}{q_E+1}$, which implies that

$$Pr(\neg E_1 \wedge \neg E_2) = \delta^{q_E}(1-\delta) \geq \frac{1}{e(1+q_E)} \quad (10)$$

Thus, we can conclude that $\mathcal{F}$ has an advantage $\epsilon_{\mathcal{F}} \geq \frac{\epsilon}{e(1+q_E)}$. □

Lemma 2. *Let H_3 be a random oracle from $\mathbb{Z}_q^*$ to $\{0,1\}^n$. Then there is an algorithm $\mathcal{F}$ that solves the BDH problem with advantage*

$$\epsilon_B \geq \frac{2\epsilon}{q_{H_3}} \quad (11)$$

Proof. Given $(g, P_1, P_2, P_3) = (g, ag, bg, cg)$, $\mathcal{F}$ sets $g_1 = P_1$ and $p_{ID} = P_2$. $\mathcal{A}$ then issues H_3 queries for h_i.

H_3 queries: Suppose that $\mathcal{F}$ keeps a list L_2 for $\langle h_i, H_3(h_i)\rangle$. If $H_3(h_i)$ exists in L_2, return that value. Otherwise, $\mathcal{F}$ randomly picks $v \xleftarrow{R} \{0,1\}^n$ and sets $H_i(h_i) = v$.
Challenge: $\mathcal{A}$ outputs M_0 and M_1. $\mathcal{F}$ randomly picks $Y \xleftarrow{R} \{0,1\}^n$ and defines $C = \langle P_3, Y\rangle$. $\mathcal{F}$ gives C to $\mathcal{A}$. Note that by definition, the decryption is

$$Y \oplus H_3(Y\|\hat{e}(Y, d_{ID})) = Y \oplus H_3(Y\|D)$$

where $D = \hat{e}(Y, d_{ID})$.
Guess: $\mathcal{A}$ outputs $u' \xleftarrow{R} \{0,1\}$. $\mathcal{F}$ randomly picks $\langle h_j, H_3(h_j)\rangle \xleftarrow{R} L_2$ and outputs h_j. Note that with Y and h_j, D can be computed because $Y\|h_j = D$, where D is the solution to the BDH problem.

Let E_H be the event that $\mathcal{A}$ issues H_3 queries for $H_3(Y\|D)$, then from [25] we know that $Pr[E_H] \geq \epsilon$ and thus $\epsilon_B \geq \frac{2\epsilon}{q_{H_3}}$. □

Theorem 1. *Suppose H_1 and H_3 are random oracles, $\mathcal{A}$ has advantage ϵ against IBE within running time t. $\mathcal{A}$ also makes q_E private key extraction queries and q_{H_3} H_3 queries. Then there exists polynomial algorithm $\mathcal{F}$ that solves the BDH problem with advantage*

$$\epsilon' \geq \frac{2\epsilon}{e(1+q_E)q_{H_3}} \quad (12)$$

within a time $t' < t + (q_{H_1} + q_E + q_{H_3})t_m$ where t_m is the time to compute a scalar multiplication in $\mathbb{G}_1^*$.

Theorem 1 follows Lemmas 1 and 2 directly.

5.3. Identity-based signature security

Security models

Definition 2 ((Strongly existentially unforgeable identity-based signature scheme under chosen-message attacks) [16]). *If no probabilistic polynomial time adversary has a non-negligible advantage in this game:*

(1) The challenger runs the setup algorithm to generate the system's parameters and sends them to the adversary.
(2) The adversary $\mathcal{A}$ performs a series of queries:

- *Key extraction queries: $\mathcal{A}$ produces an identity ID_i and receives the private key d_i.*
- *Signature queries: $\mathcal{A}$ produces a message M and an identity ID_i and receives a signature on M that was generated by the signature oracle using the private key corresponding to the identity ID_i (i.e., d_i).*
- *After a polynomial number of queries, $\mathcal{A}$ produces a tuple (ID, M, σ) made of an identity ID, whose corresponding private key was never asked during stage 2, and a message signature pair (M, σ) such that σ was not returned by the signature oracle on the input (M, ID) during stage 2 for the identity ID.*

$\mathcal{A}$ wins the game if the forged signature can be verified when the verification algorithms run on the tuple (ID, M, σ). The forger's advantage is defined to be its probability of producing a forgery taken over the number of coin-flipping of the challenger and $\mathcal{A}$.

Security analysis

Theorem 2. *Let H_1 and H_2 be random oracles, $\mathcal{A}$ has advantage ϵ against IBS in running time t. $\mathcal{A}$ also makes*

q_E private key extraction queries, q_{H_2} H_2 queries, and q_S signature queries. Then there is an algorithm $\mathcal{F}$ that solves the CDH problem with advantage

$$\epsilon_C \geq \frac{\epsilon - q_S(q_{H_2} + q_S)/q}{e(q_E + 1)} \tag{13}$$

within running time $t' < t + (q_{H_1} + q_E + q_{H_2} + 2q_S)t_m + (q_S + 1)t_{mm}$, where t_m is the running time for a scalar multiplication in $\mathbb{G}_1^*$ and t_{mm} is the running time for a multi-exponentiation in $\mathbb{G}_1^*$.

Proof. Let $P_1 = ag$ and $P_2 = bg$ be the input of the CDH problem. Given (P_1, P_2). First, $\mathcal{F}$ initializes $g_1 = P_1$ as the domain secret. From the perspective of the adversary, the distribution of g_1 secret is identical to the real one (i.e., $g_1 = sg$). Let T be a random variable following Bernoulli distribution, that is, $T \in \{0, 1\}$ with probability δ of being 0 and probability $1 - \delta$ of being 1. Then, $\mathcal{F}$ issues a series of queries as stated in the following:

(1) H_1 *queries*: Suppose the adversary issues a query for an identity ID_i. $\mathcal{F}$ first picks a random number $u_i \xleftarrow{R} \mathbb{Z}_q^*$ and decides the public key based on the outcome of T, such that

$$p_i = \begin{cases} u_i P_2, & T = 1, \\ u_i g, & T = 0 \end{cases} \tag{14}$$

$\mathcal{F}$ then keeps (ID_i, u_i, T) in list L_1.

(2) *Private key queries*: For ID_i, $\mathcal{F}$ recovers u_i and T from L_1, such that

$$(ID_i, u_i, T) \leftarrow L_1$$

And the private key of ID_i is determined as

$$d_i = \begin{cases} Abort, & T = 1, \\ u_i P_1, & T = 0 \end{cases} \tag{15}$$

$T = 1$ indicates that no answer to the query. When $T = 0$, note that $d_i = ap_i$ follows the distribution of real secret key.

(3) H_2 *queries*: Assuming that $\mathcal{F}$ keeps a list L_2 that stores any previously defined $H_2(h_i)$. Given (ID_i, M, U_i), $\mathcal{F}$ first checks whether $H_2(h_i)$ has been defined (e.g., H_i). If so, the defined value will be returned. Otherwise, $\mathcal{F}$ randomly picks $v_i \xleftarrow{R} \mathbb{Z}_q^*$, and determine $H_2(h_i)$ as,

$$H_2(h_i) = \begin{cases} v_i, & \text{not defined in } L_2, \\ H_i, & \text{already defined in } L_2 \end{cases} \tag{16}$$

(4) *Signature queries*: $\mathcal{A}$ randomly chooses $\mu_i \xleftarrow{R} \mathbb{Z}_q^*$ and $v_i \xleftarrow{R} \mathbb{Z}_q^*$. Then set $U_i = \mu_i g$ and $V_i = v_i g_1$. Define $H_2(h_i) = (u_i P_2)^{-1}(v_i g - U_i) \in \mathbb{Z}_q^*$. The pair $\sigma_i = \langle U_i, V_i \rangle$ appears as a valid signature.

$$\sigma_i = \begin{cases} \langle U_i, V_i \rangle, & H_2(h_i) \text{ not defined in } L_2, \\ Abort, & H_2(h_i) \text{ defined in } L_2 \end{cases} \tag{17}$$

(5) *Signature forgery*: Given a message M and an identity ID, $\mathcal{A}$ forges a signature $\langle U, V \rangle$. $\mathcal{F}$ recovers

$$(ID, u, T) \leftarrow L_1$$

If $T = 0$ then abort. Otherwise (i.e., $T = 1$), the list L_2 must contain an entry (ID, M, U, v) with overwhelming probability. Because $H_2(h_{ID})$ has been defined as v, if $\mathcal{A}$ succeeds in the game then $\mathcal{F}$ knows that

$$\hat{e}(g, V) = \hat{e}(g_1, p_{ID} h_{ID} + U)$$

With $h_{ID} = v$, $p_{ID} = uP_2$, where u and v are known, then $\mathcal{F}$ also finds that

$$\begin{aligned} \hat{e}(g, V) &= \hat{e}(g_1, p_{ID} h_1 + U) \\ &= \hat{e}(g_1, p_{ID} h_{ID}) \hat{e}(g_1, U) \\ &= \hat{e}(P_1, uP_2 v) \hat{e}(P_1, U) \\ \Rightarrow \hat{e}(g, V - vP_1) &= \hat{e}(P_1, vuP_2) \end{aligned}$$

And $(vu)^{-1}(V - vP_1)$ is the solution to the CDH instance (P_1, P_2).

From Lemma 1, we know that the probability of $\mathcal{F}$ not aborting in the process of key extraction query is at least $1/e(q_E + 1)$. Moreover, $\mathcal{F}$ aborts in the process of signature queries is at most $q_S(q_{H_2} + q_S)/q$ because of conflict on H_2, where q is the size of $\mathbb{G}_1$. Overall, $\mathcal{F}$ has an advantage at least $\epsilon' \geq \frac{\epsilon - q_S(q_{H_2} + q_S)/q}{e(q_E + 1)}$. That completes the proof. □

6. PERFORMANCE ANALYSIS OF THE PROPOSED SCHEMES

6.1. Analysis

Performance of the proposed schemes is based on the number of operations and the efficiency of each type of operations. Table II lists the number of operations of each algorithm. Among them, *mul* indicates standard multiplication in $\mathbb{G}_1$. Because the addition in $\mathbb{G}_1$ and XOR

Table II. Computation complexity.

	# of $\hat{e}$	# of *mul*	# of H_1	# of H_2	# of H_3
Signcrypt	1	5	1	2	1
Decrypt	1	0	0	0	1
Sign	0	3	1	0	0
Verify	3	1	1	1	0

are simple and efficient operations, they are not listed in the table.

Hash functions can be computed efficiently in general. In practice, H_2 and H_3 are easy to find. However, it is hard to build $H_1 : \{0,1\} \rightarrow \mathbb{G}_1^*$. In the analysis, we relax H_1 into two steps.

(1) $H_1 : \{0,1\}^* \rightarrow I \subseteq \{0,1\}^*$;
(2) $H_1' : I \rightarrow \mathbb{G}_1^*$

In step 1, I is a finite set, H_1' is an encoding function that is computable. Note that after the relaxation, the public key for a give *ID* and *time* is $p_{ID} = H_1'(H_1(ID||time))$. In the proposed IBSC scheme, public keys can be computed at the beginning of each session and cached for the entire session. Therefore, the relaxation of H_1 does not introduce more computation cost in reality. Because of that, performance of the IBSC and IBS schemes will be considered efficient if bilinear pairing $\hat{e}$ and multiplication in $\mathbb{G}_1$ can be computed efficiently. Because the Weil pairing can be performed efficiently using Miller's algorithm [26], the bilinear map $\hat{e}$ can be performed efficiently as well.

To analyze the performance of the IBSC scheme, we apply two bilinear pairing functions, that is, *modified Weil pairing* and *Tate pairing* over supersingular elliptic curve $E : \{y^2 = x^3 + 1 | x, y \in \mathbb{F}_p\}$. We first construct $\mathbb{G}_1$. Let p be a prime number s.t. $p \equiv 2 \mod 3$ and $p = aq - 1$ for some prime q and positive integer a. Then $\mathbb{G}_1$ is the subgroup of order q of $\mathbb{F}_{p^2}^*$. CCH problem is hard in the group $\mathbb{G}_1$ [25,27]. However, it is worth mentioning that *decisional Diffie–Hellman problem* is an easy one for bilinear map $\hat{e}$. This is because with given $P, aP, bP, cP \in \mathbb{G}_1$, $\forall a, b, c \in \mathbb{Z}_q^*$, we can easily check if $c \equiv ab \mod q$ by comparing $\hat{e}(aP, bP)$ with $\hat{e}(P, cP)$.

The Weil pairing e has the properties of bilinearity and computability; however, it does not have non-degeneracy. Therefore, we adopt a modified Weil pairing $\hat{e} : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ s.t. $\hat{e}(P, Q) = e(P, \phi(Q))$, where ϕ is an automorphism on the group of points of supersingular elliptic curve $E : \{y^2 = x^3 + 1 | x, y \in \mathbb{F}_p\}$, that is, $\phi(x, y) = (\xi x, y)$, where ζ is a primitive cube root of unity in $\mathbb{F}_p$. Thus, $y^2 = (\zeta x)^3 + 1 = \zeta^3 x^3 + 1 = x^3 + 1 \Rightarrow \phi(P_1) + \phi(P_2) = \phi(P_1 + P_2), \forall P_1, P_2 \in \mathbb{G}_1$. The bilinear map $\hat{e}$ is calculated as a Weil pairing with an additional standard multiplication on the curve E. According to [25], $\hat{e}$ is believed to satisfy the BDH problem. However, computing discrete logarithm in $\mathbb{F}_p^*$ is sufficient for computing discrete logarithm in $\mathbb{G}_1$. Therefore, in order to make it sufficiently hard in practice, q needs to be at least 512-bit long.

6.2. Numerical results

We evaluate the proposed identity-based schemes with modified Weil pairing $\hat{e}$ using Mathematica 10.0 with a computer equipped with an Intel Core i5-2400 @ 3.1 GHz and 12 GB RAM. We first show the computational cost of each operation. Because H_1, H_2, and H_3 do not have much difference in computational time and the added encoding function H_1' is more efficient than H_1, the hash functions are excluded from the performance analysis.

First, we evaluate the computational time for bilinear pairing $\hat{e}$. Two sets of evaluation are given, that is, for $q = 256$ bits and $q = 385$ bits. Each evaluation is the average value from 10 000 calculations. With $q = 256$ bits, one $\hat{e}$ takes about 7.44 ms. With $q = 385$ bits, one $\hat{e}$ takes about 13.25 ms. We then evaluate the computational time of standard multiplication over $\mathbb{G}_1$ (i.e., $k_p P \in \mathbb{G}_1$). The computational time of $k_p P \in \mathbb{G}_1$ mainly depends on the size of k_p (assuming $q = 256$ bits). The computational time of each evaluation is averaged from 10 000 calculations. With $k_p = 128/256/512$ bits, a standard multiplication operation takes about 3.25/6.43/12.29 ms. Note that in the proposed IBSC, k_p is the output of some hash functions, therefore, k_p usually is 256 bits or 512 bits, where the computation is efficient. The evaluation results are summarized in Table III.

Based on the evaluations we have for each operation, we then show the total operational time for each algorithm. In practice, public keys are computed once and cached for the entire session. Computational time of each algorithm is listed in Table IV. It is shown that the proposed IBSC performs efficiently for delay tolerable and even near real-time data transmission, for example, metering data transmission. However, for real-time monitoring data, for example, PMU data, identity-based schemes alone may not be a good solution. Without sufficient computational resources, faster security protocols and schemes are recommended, for instance, traditional symmetric ciphers. The proposed IBSC can be applied for initial authentication and key distribution of the chosen symmetric ciphers.

Table III. Computational time for each operation.

Bilinear pairing $\hat{e}$	
$q = 256$ bits	$q = 385$ bits
7.44 ms	13.25 ms
Standard multiplication	
$k_p = 256$ bits	$k_p = 512$ bits
6.43 ms	12.29 ms

Table IV. Computational time of each algorithm.

	$q = 256$ bits $k_p = 256$ bits	$q = 256$ bits $k_p = 512$ bits
Signcrypt	39.59 ms	68.89 ms
Decrypt	7.44 ms	7.44 ms
Sign	19.29 ms	36.87 ms
Verify	28.75 ms	34.61 ms
	$q = 385$ bits $q = 256$ bits	$q = 385$ bits $q = 512$ bits
Signcrypt	45.4 ms	74.7 ms
Decrypt	13.25 ms	13.25 ms
Sign	19.29 ms	36.87 ms
Verify	46.18 ms	52.04 ms

7. CONCLUSION

In this paper, we proposed a big data-driven, cloud-based ICT framework for smart grid communication infrastructure. Taking into consideration the security requirements of each message, we proposed an ID-based signcryption security scheme to secure the transmissions in the ICT framework. The proposed IBSC scheme performs simultaneously the functions of encryption and digital signature. Therefore, confidentiality, non-repudiation as well as data integrity are provided. The proposed IBSC scheme was also reduced to an ID-based digital signature scheme. To further enhance the computational performance, symmetric ciphers are introduced to the IBSC. In addition, signing right delegation from one LCC to another (or a few) LCC is achieved by identity-based schemes. The security of the proposed IBSC is studied. The numerical results showed that the proposed IBSC scheme is able to perform efficiently with security guarantee in the CPS of smart grid. In the future, we will focus on the researches of big data analytics in implementing the ICT framework in smart grid.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation under grants CNS-1423348 and CNS-1423408.

REFERENCES

1. Zhou J, Hu RQ, Qian Y. Scalable distributed communication architectures to support advanced metering infrastructure in smart grid. *IEEE Transactions on Parallel and Distributed Systems* 2012; **23**(9): 1632–1642.
2. Yan Y, Qian Y, Sharif H, Tipper D. A survey on smart grid communication infrastructures: motivations, requirements and challenges. *IEEE Communications Surveys Tutorials* 2013; **15**(1): 5–20.
3. Ye F, Qian Y, Hu RQ. Energy efficient self-sustaining wireless neighborhood area network design for smart grid. *IEEE Transactions on Smart Grid* 2015; **6** (1): 220–229.
4. Ye F, Qian Y, Hu RQ, Das SK. Reliable energy-efficient uplink transmission for neighborhood area networks in smart grid. *IEEE Transactions on Smart Grid* 2015; **6**(5): 2179–2188.
5. Maharjan S, Zhu Q, Zhang Y, Gjessing S, Basar T. Dependable demand response management in the smart grid: a Stackelberg game approach. *IEEE Transactions on Smart Grid* 2013; **4**(1): 120–132.
6. Ye F, Qian Y, Hu RQ. A real-time information based demand-side management system in smart grid. *IEEE Transactions on Parallel and Distributed Systems* 2016; **27**(2): 329–339.
7. Ye F, Qian Y, Hu RQ. Incentive load scheduling schemes for PHEV battery exchange stations in smart grid. *Systems Journal, IEEE* 2015; **PP**(99): 1–09.
8. Yan Y, Qian Y, Sharif H, Tipper D. A survey on cyber security for smart grid communications. *IEEE Communications Surveys Tutorials* 2012; **14**(4): 998–1010.
9. Siddiqui F, Zeadally S, Alcaraz C, Galvao S. Smart grid privacy: Issues and solutions. *2012 21st International Conference on Computer Communications and Networks (ICCCN)*, Munich, Germany, 2012; 1–5.
10. Baek J, Vu QH, Liu JK, Huang X, Xiang Y. A Secure Cloud Computing Based Framework for Big Data Information Management of Smart Grid. in *IEEE Transactions on Cloud Computing* 2015; **3** (2): 233–244.
11. Chuang C-L, Wang Y-C, Lee C-H, Liu M-Y, Hsiao Y-T, Jiang J-A. An adaptive routing algorithm over packet switching networks for operation monitoring of power transmission systems. *IEEE Transactions on Power Delivery* 2010; **25**(2): 882–890.
12. Rusitschka S, Eger K, Gerdes C. Smart grid data cloud: a model for utilizing cloud computing in the smart grid domain. *2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Gaithersburg, MD, USA, 2010; 483–488.
13. Zheng L, Hu Y, Yang C. Design and research on private cloud computing architecture to support smart grid. *2011 International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 2011; 159–161.
14. Bitzer B, Gebretsadik ES. Cloud computing framework for smart grid applications. *2013 48th International Universities' Power Engineering Conference (UPEC)*, 2013; 1–5.
15. Waters B. Efficient identity-based encryption without random oracles. *Advances in Cryptology–EUROCRYPT 2005*, Aarhus, Denmark, 2005; 114–127.
16. Libert B, Quisquater J-J. The exact security of an identity based signature and its applications. *IACR Cryptology ePrint Archive* 2004; **2004**: 1–19.
17. Ye F, Qian Y, Hu RQ. HIBaSS: hierarchical identity-based signature scheme for ami downlink transmission. *Security and Communication Networks* 2015; **8** (16): 2901–2908, DOI 10.1002/sec.1217.
18. Ye F, Qian Y, Hu RQ. A security protocol for advanced metering infrastructure in smart grid. *Global Communications Conference (GLOBECOM), 2014 IEEE*, Austin, TX, USA, 2014; 649–654.
19. Simmhan Y, Aman S, Kumbhare A, Liu R, Stevens S, Zhou Q, Prasanna V. Cloud-based software plat-

form for big data analytics in smart grids. *Computing in Science Engineering* 2013; **15**(4): 38–47.
20. Ukil A, Zivanovic R. Automated analysis of power systems disturbance records: smart grid big data perspective. *Innovative Smart Grid Technologies - Asia (ISGT Asia), 2014 IEEE*, Kuala Lumpur, Malaysia, 2014; 126–131.
21. Kanabar M, Adamiak MG, Rodrigues J. Optimizing wide area measurement system architectures with advancements in phasor data concentrators (PDCS). *Power and energy society general meeting (PES), 2013 IEEE*, Vancouver, Canada, 2013; 1–5.
22. Hu H, Wen Y, Chua T-S, Li X. Toward scalable systems for big data analytics: a technology tutorial. *Access, IEEE* 2014; **2**: 652–687.
23. Tan W, Blake MB, Saleh I, Dustdar S. Social-network-sourced big data analytics. *Internet Computing, IEEE* 2013; **17**(5): 62–69.
24. Ji C, Li Y, Qiu W, Awada U, Li K. Big data processing in cloud computing environments. *2012 12th International Symposium on Pervasive Systems, Algorithms and Networks (ISPAN)*, San Marcos, TX, USA, 2012; 17–23.
25. Boneh D, Franklin M. Identity-based encryption from the weil pairing. *SIAM Journal on Computing* 2003; **32**(3): 586–615.
26. Park CM, Kim MH, Yung M. A remark on implementing the weil pairing. *Information Security and Cryptology*, Springer Berlin Heidelberg, 2005; 313–323. ISBN: 978-3-540-30855-3, DOI: 10.1007/11599548_27.
27. Joux A, Nguyen K. Separating decision Diffie–Hellman from computational Diffie–Hellman in cryptographic groups. *Journal of Cryptology* 2003; **16**(4): 239–247.