Yu-fei WANG, Jian-ming FU*, Yan LIN, Wei-xia LI

Information Flow Integrity of ECPS based on Domain Partition

Abstract: A smart grid is faced with many security threats, such as false data injection attacks, replay attacks and delay attacks. In this paper, we have constructed an information flow model which abstracts the ECPS workflow as the information flows among nodes, and nodes with a similar function is grouped into the same domain. On account of the above model, we define and clarify the threat model in ECPS. Then, this paper proposes an information flow integrity assurance mechanism, which includes secure communication framework and key management. Encryption algorithms and digital signatures have been used in the mechanism, which share the public-private key pair in the same domain to decrease the complexity of key management. Finally, we analyzed and verified the effectiveness of our mechanism by using a Ukraine power failure event on the transformer substation. After combining the information flow model and the mechanism we proposed, the above three kinds of attacks can be countered and the information such as data, command can be integrated and confidential during the process of information collection and transmission.

Keywords: smart grid; cryptography; information flow; integrity; key management

1 Introduction

Compared to traditional power grids, smart grids allows a great number of users to access it. And it can monitor energy flow and consumption in real time, which provides support for the efficient use of energy, zone pricing, fault isolation and decision making. The implementation of smart grid is based on Electrical Cyber-Physical System(ECPS), and the prototype of ECPS is a Cyber-Physical System(CPS) [1], which leverages the interaction between a great number of calculating units and physical units, and the transmission of control flow to achieve a real-time monitoring for a physical system.

Smart grids have achieved the industry innovation required by the current age, but also face some security challenges. On one hand, the boundary of smart grid has been widened by the great number of accessed users and distributed power, which increases as the attack surfaces; on the other hand, the critical modules in smart
grid often use IP networks to connect each other, and the communication protocol is becoming public and standardized, so there are common threats in these modules as well as in the information system. For example, attackers can tamper with, counterfeit data and issue false commands to destroy the availability, integrity, dependability and controllability of the modules and devices [2]. Smart grids have been exploited by malware in recent years. Stuxnet [3] exploits the virus in a mobile storage medium to diffuse malware to smart grid, which is the malware that first attacked on the infrastructure in industry. Blackenery [4] is another malware, which can release back doors and special module to issue false commands and delete data on disks, and blackenergy has generated serious threats. For example, the Ukraine power failure event caused power failure in more than half of the region.

Encryption and identity authentication are effective ways to ensure the security of the cyber system in a smart grid [5-8]. However, the key management is a complex problem, including key replacement, key distribution, key security, key operation cost and so on. In order to reduce the complexity of key management, Kim et al. [9] and Wu et al. [10] proposed some ways to decrease the key operation cost and increase the security of key respectively.

All of these work provide some cryptography methods to improve the security of ECPS, but they do not provide a systematical model for ECPS, and they do not consider key management systematically.

In this paper, we have constructed an information flow model of ECPS based on domain partition, such as data collection domain, data processing domain, decision domain, and control domain. According to this model, we have proposed an information flow integrity mechanism based on encryption and digital signature, and each node in same domain shares the same public-private key pair, so the complexity of the key management in smart grid is decreased. This mechanism can counter against false data injection attacks, replay attacks, delay attacks, and ensure that the data and the command are integrated and confidential during the procedure of transmission.

To measure the effectiveness of our information flow integrity mechanism, the Ukraine power failure event on transformer substation is analyzed, and the result shows that our mechanism can effectively defend from these kind of attacks.

The rest of this paper is organized as follows. The next section briefly states related work, and our information flow model has been presented based on domain partition in ECPS in section III. Section IV introduces the information flow integrity mechanism based on encryption and identity authentication. Section V illustrates a case study for our mechanism. Finally, we have given a conclusion and a future direction for work.
2 Related Work

A smart grid gives a new mission to the power grid, and its informatization, intelligence, automation propel the revolution of the smart grid infrastructure. At the same time, the security problems in smart grids attract many researchers.

Many warnings concerning the security of smart grid are appearing [2,5,11-13]. And ECPS [2,11,12] is a complex and heterogeneous system, which is formed by the coordination between a cyber system and a physical system. Both of the two systems interdepend with each other, so it is more complex to analyze the security and dependability of ECPS than a single system. The open technologies in ECPS rely on computer systems and communication systems, so there are information security threats to it. How to ensure the availability, integrity and confidentiality of the information flow in ECPS is a key issue. Lots of data are generated during the interactions of a cyber system and a physical system [13], such as real-time measurements, historical data, external events and decisions, and it is also crucial to ensure these data are transmitted efficiently and securely. If internal staff use the picture steganography to carry out illegal information, we should detect the abnormal image [20,21]. Liu Xueyan et al. [5] analyzed the threats in a smart grid from devices, network and data, and proposed some techniques can be applied to smart grid, for instance, intrusion detection, access control and industrial firewall. They also pointed out that key management in a smart grid is a challenge. But there is no solution to the characteristics of ECPS.

To improve the security in a smart grid, researchers have proposed lots of methods. Lu et al. [8] proposed that Advanced Metering Infrastructure(AMI) is a key module in smart grid, and analyzed some common attacks on it, including packet sniffing, data interception, data falsification, data tampering, DDoS attack and so on. Meanwhile, they used a light-weight symmetric encryption algorithm Blom to ensure the integrity and confidentiality of the communication data, which can be used to counter false data injection, replay attack and man-in-the-middle attack, but ZIGBEE-based wireless networks and other protocol frame structures have not yet been proposed. Mo et al. [14] analyzed sources of threats in a smart grid, such as price information, commands, measurements and software. Based on system theory, they built a model for smart grid and its threats, then they used encryption algorithm and identity authentication to achieve confidentiality and entity authentication respectively. However, the physical world is modeled with approximations and is subject to noise, which can result in a deviation of any model to the reality. Therefore, system-theoretic approaches are nondeterministic as compared to information security. Liu et al. [6] described a dynamic approach which dynamically generates a key for sender and receiver in wireless communication, and it increases the difficulty to launch an attack. However, it also increases the complexity of key management. Mike et al. [7] extended the traditional Byzantine fault model to prevent collaborative
exploits. The traditional Byzantine fault model just considers the potential attacks in CPS, but it does not consider that the attacker can intercept and tamper with the data.

Key management is a crucial problem in a smart grid. Kim et al. [9] leveraged a binary tree to manage key search among different nodes, which achieves the security for unicast, multicast and broadcast communication. By using binary tree, the time of querying a shared key decreased from $O(n^2)$ to $O(\log_2(n))$, but this method relies on a third party for identity authentication or key generation. It might cause additional equipment cost and communication traffic. Wu et al. [10] integrated symmetric encryption algorithm and elliptic curve cryptography to dynamically generate key to counter man-in-the-middle attack. In their scheme, they combine both of PKI and a third trusted anchor, which will essentially increase the complication for the smart grid because their protocol at least needs two different kinds of servers for PKI and the trust anchor respectively. Besides, this scheme is not secure against the man-in-the-middle attack.

Compared to some existing work, our work not only counters against false data injection attacks, replay attacks and delay attacks, but also decreases the number of public-private key pairs needed in a smart grid, which is helpful to reduce the complexity of key management.

3 Information Flow And Threat Model

A smart grid is composed of four parts: infrastructure layer, measurement layer, information processing layer and decision making layer. The first one locates in the physical system and the last three constitute the cyber system in smart grid.

3.1 Information flow model

The work flow of ECPS can be described as follows: physical sensors collect real-time data from electric equipment components, and then transmit these data to its upper-level component; the upper-level component process these data, and make a decision, and then the decision is transmitted to the control component; finally, the control component will issue the decision to achieve energy management. We abstract the work flow of ECPS into information flow between nodes, and we put nodes which have the similar functionality into a domain, as shown in Figure 1. The information collection domain and control domain form the measurement layer, which achieves the interaction between cyber system and physical system in smart grid; the information processing layer is constituted by data processing domain, and the decision making domain corresponds to the decision making layer.
In order to describe the data flow of ECPS more accurately, we give some definitions as follows:

Definition 1 (ECPS-IFG). A ECPS-IFG (ECPS Information Flow Graph) = <D, E, δ >. D is the set of domains formed the ECPS; E is the set of data flows between different domains, and δ is the data flow function.

Definition 2 (Domain-IFG, D-IFG). A Domain Information Flow Graph in ECPS can be described as Domain-IFG = <N, E, F >. N is the set of nodes a domain has, and there are two kinds of nodes in a domain: boundary node N_b and internal node N_i. The responsibility of N_b is to transmit data, and N_i is responsible to process data. When there is only one node in a domain, the type of this node is boundary node beside internal node; E is the set of data flows in the domain, and F is the data flow function in the domain.

Given a ECPS-IFG = <D, E, δ >, D_i and D_j are two different domains in the system, the following relations are true:

1. the number of nodes in D_i and D_j are more than one.
   \[ \exists n_{hk} \in N_b \in D_i \text{ makes } \exists E : n_{hk} \to n_{jr} \in N_b \in D_j \text{ That is to say, there will be an information flow E from boundary node } n_{hk} \text{ to a boundary node } n_{jr} \text{ in another domain } D_j \]
   \[ \exists n_{ij} \in N_i \in D_i \text{ makes } \exists E : n_{ij} \to n_{tv} \in N_i \in D_i \]
   That is to say in domain D_i, there will be information flow E from internal node n_{ij} to a node n_{tv}, and n_{tv} can be a boundary node or an internal node.

2. the number of nodes in D_i is equal to one.
   For n \in D_i, \exists E : n \to n_{tv} \in D_i. When the number of nodes in D_i is more than one, n_{tv} \in N_b, that is to say when domain D_i only has one node, in the function of \delta, the information flow will be transmitted to the node in another domain D_j, and if the number of nodes in D_j is more than one, this node must be boundary node.

Figure 1. Information flow model of ECPS
As shown in Figure 1, the cyber system is composed of four domains: data collection domain \( D_1 \), data processing domain \( D_2 \), decision making domain \( D_3 \) and control domain \( D_4 \); the physical system has three parts: data emission domain \( P_1 \), data processing domain \( P_2 \) and command receiving domain \( P_3 \). And there is only one node in domain \( D_1 \), \( D_4 \), \( P_1 \) and \( P_3 \), the functionality of each node and the information flow in ECPS are described as follows:

1. Node \( P_1 \) in data emission domain, gets the information from \( P_2 \) and sent it to \( D_1 \).
2. Node \( D_1 \) in data collection domain, collects the real-time data from physical system and transmits the data to the boundary node \( D_{21} \) in data processing domain.
3. Boundary node \( D_{21} \) in data processing domain, just transmits the data to the internal node \( D_{23} \).
4. Internal node \( D_{23} \) in data processing domain, stores and computes the data it received and then transmits it to boundary node \( D_{22} \).
5. Boundary node \( D_{31} \) in data processing domain, transmits data to the boundary node \( D_{32} \) in decision making domain.
6. Boundary node \( D_{31} \) in decision making domain, transmits data to internal node \( D_{33} \).
7. Internal node \( D_{33} \) in decision making domain, analyzes the data it received and make some decisions on it. Finally, it will transmit the decision to the boundary node \( D_{32} \).
8. Boundary node \( D_{32} \) in decision making domain, transmits the decision to control domain.
9. Node \( D_4 \) in control domain issues the decision to smart devices.
10. Node \( P_3 \) in commands receiving domain, gets the information from \( D_4 \) and sent it to \( P_2 \).
11. Node \( P_2 \) in data processing domain, gets the commands from \( P_3 \) and processes it, then tells physical equipment what to do and how to work.

### 3.2 Threat Model

Some attacks in smart grid can be directly reflected to our information flow model, such as data interception, data tampering and false data injection. According to the information flow model, we define the threat model in ECPS as follows.

1. For any domains in ECPS, attackers can inject false data in it.
2. For the information flow from domain \( D_i \) to domain \( D_j \), attackers can intercept the datagram transmitted. That is to say, the datagram transmitted between two different domains can be intercepted.
3. Attackers can use the false data and the data intercepted to launch a replay attack or a delay attack to any domains.
Figure 2 shows the situations when false data injection attacks, replay attacks and delay attacks happen.

![Attack model in ECPS](image)

**Figure 2.** Attack model in ECPS

Attackers send a false datagram to the cyber system, or intercept the datagram and then tamper with it, so ECPS cannot perceive the actual state of the physical devices and decision making domain will make an improper or wrong decision. As shown in Figure 2, the attacker injects false data to domain $D_2$, which makes the boundary node $D_{22}$ transmit the false data to the decision making domain. Hence, the wrong decision is issued to the smart device and device disturbance occurs.

A replay attack means attackers intercept the datagram transmitted between two different domains and then send it repeatedly to bypass the identity authentication, even to launch a Denial of Service (DoS) attack. As shown in Figure 2, the attacker intercepts the datagram transmitted from domain $D_2$ to domain $D_3$.

When attackers delay the datagram arriving time between two different domains is called delay attack, which will make the receiver get the obsolete datagram. Shown in Figure 2, when the attacker intercepts the datagram from $D_2$ to $D_3$, he does not send it immediately, but sends it after some time.

On 23rd December, 2015, the Ukraine power sector was attacked, which caused power failure among more than half of the region, and its attack flow is described in Figure 3. Firstly, an email with malware was sent to users in the power sector. When a user clicked this email, the Trojan was run, and a SSH back door and a killdisk module were installed, then with this back door the attacker could issues command to smart devices. In addition, killdisk can delete system data to delay the system recovery time.
Information Flow Integrity of ECPS based on Domain Partition

Figure 3. Attack flow of Ukraine power sector

From the Ukraine power failure event, we can find that a smart grid now will be under great threat, and there will be heavy losses when attackers issue false command to the electric devices. So it is urgent to ensure the security and dependability of ECPS [4,15,16].

4 Information Flow Integrity Mechanism

According to the information flow integrity model and threat model introduced in section III, we propose an information flow integrity mechanism, which includes secure communication framework and key management. This mechanism can ensure that the data transmitted in the ECPS can only be shared by authorized nodes, and it is not counterfeited and tampered with by attackers.

4.1 Secure communication framework

As introduced in section III-B, false data injection attack has two parts. One is data interception and modification, and the other one is to inject false data into a domain directly. In order to counter it, we need to use cryptography method to ensure data cannot be divulged and modified during transmission, and to ensure the sender is authorized by using digital signatures.

False data injection attacks can be mitigated by using cryptography methods and digital signatures, but the attacker can intercept the data which is encrypted and signed, and send it repeatedly to cause a DoS attack. So a timestamp is added at the process of computing a digital signature, and the receiver needs to check whether the timestamp is valid after receiving a datagram.

There are two kinds of cryptography methods, symmetric encryption and asymmetric encryption. We use symmetric encryption to achieve confidentiality for data and commands. The data processing domain and decision making domains can use cryptography methods that have a high level of security [22,23], such as AES. For control domains we can use some lightweight encryption methods, such as MIBS [17],
SMS4 [18] and PRESENT [19]. RSA is the standard algorithm used in digital signature, so we use it to achieve identity authentication. The secure communication framework is shown in Figure 4, and the functionality of each node is shown in Table 1.

![ECPS secure communication framework](image)

**Figure 4.** ECPS secure communication framework

<table>
<thead>
<tr>
<th>Node</th>
<th>Functionality</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1$</td>
<td>collect real-time data from electric devices, do encryption and digital signature on the data and then send it to boundary node $D_{21}$</td>
</tr>
<tr>
<td>$D_{21}$</td>
<td>transmit data to node $D_{23}$</td>
</tr>
<tr>
<td>$D_{23}$</td>
<td>signature verification and decryption; then store, compute the data, encrypt and sign the data been operated and transmit the datagram to node $D_{22}$</td>
</tr>
<tr>
<td>$D_{22}$</td>
<td>transmit datagram to node $D_{31}$</td>
</tr>
<tr>
<td>$D_{31}$</td>
<td>transmit datagram to node $D_{33}$</td>
</tr>
<tr>
<td>$D_{33}$</td>
<td>signature Verification and decryption; then make decision, encrypt and sign the decision and transmit the datagram to node $D_{32}$</td>
</tr>
<tr>
<td>$D_{32}$</td>
<td>transmit datagram to node $D_4$</td>
</tr>
<tr>
<td>$D_4$</td>
<td>signature Verification and decryption; issue decision to smart devices</td>
</tr>
</tbody>
</table>

**4.2 Key management**

Smart grids are a large-scale and heterogeneous network, and the large number of nodes and their connections in it make a great challenge to the encryption key and identity key management. If every node in the information flow model has its own
public-private key pair, the number of keys for ECPS will be very large, and the key
management will be very complex. We propose sharing the public-private key pair
in a domain to decrease the number of keys, which reduces the complexity of key
management. The key management is shown in Figure 5, the responsibility of KDC
(Key Distribution Center) is to distribute and manage encryption key, identity key is
distributed and managed by CA (Certification Authority). As shown in Figure 5, nodes
in data processing domain and decision making domain share the same public-
private key pair respectively.

Combined with key management, the communication flow in ECPS is described in
Figure 6, which contains five steps.
1. KDC distributes an encryption key $K$ for every domain, and identity key $(PK_{D_i},$
$SK_{D_i})$ for every domain is distributed by CA.
2. Node $D_1$ encrypts data $m$ using encryption key $K$, then computes its hash value $h$.
   In order to defend against replay attacks and delay attacks, the digital signature
   for hash value $h$ and a timestamp $T$ is computed. Finally, node $D_1$ transmits the
   encrypted data, digital signature and its public key certificate to the boundary
   node in domain $D_2$.
3. Internal node $D_{23}$ in domain $D_2$ verifies the validity of the public key certificate it
   receives and uses the public key to get the hash value $h'$ and timestamp $T'$ from
   the digital signature. Then $D_{23}$ computes the hash value using the encrypted data
   it received and compares it with $h'$. At the same time, $T'$ is checked whether it
   is closer to the current time. If both conditions are satisfied, encrypted data is
   decoded using key $K$. Finally node $D_{23}$ does some operations on data $m$ and then
   does the same steps described in step 2).
4. Node D₃ does the same operations in step 3, then it will make decision according to the information, and then transmits the encrypted decision and digital signature to the boundary node.

5. Node D₄ verifies the identity of the sender and decrypts the encrypted decision according to steps in step 3, and then delivers the decision (command) to electric devices.

Figure 6. Communication flow combined with key management
5 Case Study

In this section, we will introduce how our information flow integrity mechanism can be applied to the Ukraine power failure event and to the transformer substation.

There is a transformer substation, whose architecture is shown in Figure 7. It is composed of station level, bay level and process level. Process level is the lowest level, which controls the electric devices directly; the middle one is a bay level, which is responsible for collecting data and issuing command to smart devices; the station level is the highest one, and data processing, decision making domains are located in this level.

![Figure 7. Architecture of transformer substation](image)

When collecting data from smart devices, the measuring and control device gets its own encryption key and identity key from KDC and CA, and all the keys are stored in a trusted storage area. Then it uses the encryption key to encrypt the data collected from smart devices, and integrates the current timestamp, the hash value for encrypted data into a data group. Finally, it gets a digital signature using the identity key and the encrypted data, digital signature, and its public key certificate as a payload is transmitted to the process level.

The dispatch center verifies the validity of the public key certificate after it receives the datagram, if it is authorized, it gets the hash value and the timestamp from the digital signature, and compares it with the hash value computed from the encrypted data and checks whether the timestamp is valid. After that, decode the encrypted data and make decisions according to it, such as turning off the power network. Because attackers do not have an encryption key and a private key, they cannot launch a false data injection attack; meanwhile, there is a timestamp, so replay attacks and delay attacks cannot be successful.
When the measuring and control device receives a command, it must verify the identity of the sender. As shown in Figure 6, in the Ukraine power failure event, the attacker controlled the monitor device and let it issue a command to turn off the power network. When the measuring and control device received this command, it will verify the identity of the sender, but now the sender did not have the correct private key, so it cannot bypass the identity authentication, and the attack would fail. Similarly, the attack does not have the encryption key, and it cannot inject false command to turn off the power network.

6 Conclusion

An information flow integrity mechanism of ECPS based on domain partition can counter false data injection attack, replay attack, delay attack effectively, and it can reduce the complexity of key management in a smart grid. The data collected or the command must go through a lot of nodes, so it is easy to append digital signatures and timestamps on forward nodes. The receiver can do the inverse operations to the sender to check the digital signatures and timestamps, and detect attacks. In addition, each node can be combined with a log audit to discover traffic anomalies, to collect data anomalies and command anomalies. For example, the control domain can compute the command entropy in a time window and then rely on the difference on the command entropy to detect replay attacks, DOS attacks and false command attacks. Of course, our mechanism cannot counter against side-channel attacks during encryption or decryption. In addition, we may use a novel encryption method to counter against this attack. A smart grid is composed of electricity generation, transmission, distribution and consumption [24], and in the future, we will focus on the attack and the protection on specific object to improve the protection efficiency, and we should improve and apply this mechanism.

Acknowledgment: Supported by Science and Technology Project of China State Grid Corp. Research on the key technology of security strategy and information communication support power system (XXB17201400056).

References


