IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p-ISSN: 2278-8735
PP 51-61
www.iosrjournals.org

Prevention of Cyber Attacks in the Distribution Automation System Using Symmetric Key Encryption Algorithm

Reeja Rasheed
Department of computer science and engineering, Muslim association college of engineering, trivandrum

Abstract : *Distributed Automation System has an inherent vulnerability to cyber attacks due to its high dependency on communication and geographically widely spread terminal devices. Protecting DAS from cyber attack is a major concern of this proposed work. In this, we analyze the types of security goals in various distribution system and efforts are made to identify the most critical security goals in the distribution automation system and propose efficient ways of achieving these goals. The message authentication and integrity is far more important than any other security requirements in the distribution system applications. The proposed work addresses possible cyber threats and overcome such attacks through a secret key distribution protocol. The proposed work also contains the features such as, disconnecting intruder when an intrusion is found where the work proposed does not have this provision, Placing encrypted message in the message frame of proposed message format as additional layer of security, Decrypting message with the symmetric decryption key at the receiving end, Making use of the reserved area for cryptic flag.*

Keywords: *Cyber security, distribution automation system, power system security*

1. Introduction

Computer and network systems fall victim to many cyber attacks of different forms. To reduce the risks of cyber attacks, an organization needs to understand and assess them, make decisions about what types of barriers or protection mechanisms are necessary to defend against them, and decide where to place such mechanisms. Understanding cyber attack characteristics (threats, attack activities, state and performance impact, etc.) helps in choosing effective barriers. Understanding the assets affected by cyber attacks helps decide where to place such barriers.

Cyber attacks of different forms threaten an organization's computer and network systems. Such attacks are increasingly becoming more sophisticated and pose greater threats . Shortcomings in current detection methods come from a lack of rigorous scientific understanding of attacks. The two common methods for cyber attack detection are signature recognition and anomaly detection. Signature recognition uses a model of "bad" system behavior to detect known attacks and cannot detect novel attacks. Anomaly detection uses a model of "good" system behavior to detect novel attacks, but suffers from the generation of many false-positives (signaling an attack when none occurred). Furthermore, existing techniques are mostly developed empirically, with only test results from limited cases, rather than based on the scientific knowledge of attack and normal data. In order to enhance our ability to efficiently detect attacks, it is important to develop a clear, scientific understanding of attack and normal characteristics of computers and networks. This paper deals with the security problems in distribution automation system.

1.2 Cyber Security

Recent cyber breaches awakened the concerns about cyber Security in the SCADA systems. Recent advances in business model require the SCADA network to be connected with corporate networks. This means that the SCADA system is subject to be under the same potential cyber attacks as other corporate networks are. Moreover, the communication architecture is more relying on the open standard communication protocols. The use of the open communication protocols renders the SCADA system more vulnerable to cyber attacks. In the past few years, the security issues in the SCADA system have been analyzed and some efforts have been carried out for developing security mechanisms .The works have been focused on mostly key management schemes for cryptographic algorithm as well as transition issues for adapting security mechanisms and intrusion detection schemes.

As for the cyber attacks, the distribution system is more vulnerable in many ways. The terminal devices in the SCADA system are mostly located in restricted local area networks, while FRTUs in the distribution system are located at remote and unmanned sites in most cases, and are spread in wider area networks. As communication between the DAS server and FRTUs becomes more critical, security measures should be implemented to protect the normal control operations from any cyber threats. Recently, agent-based service-restoration algorithms in the DAS network have been proposed, and those algorithms are dependent on the security and reliability of the network .

1.3 Supervisory Control And Data Acquisition (Scada)

In the past few years the security issues in the supervisory control and data acquisition (SCADA) system have been widely investigated, and many security mechanisms have been proposed from research communities. The international standard organizations also have published several standard documents for secured SCADA systems.

The main purpose of the supervisory control and data acquisition (SCADA) system is gathering real-time data, monitoring and controlling equipments and processes in the critical infrastructure. A SCADA network provides connection between servers which resides inside a control center and control devices which are located at fields, sometimes at remote locations.

Major concern about cyber attack stems from the notion that the SCADA network is no longer an isolated network which prohibits outsiders from entering the network, nor is the specialized network based on private platforms and protocols, allowing only technical staffs with special knowledge to access to the resources. The reasons of claiming that the SCADA network is not a protected closed network is twofold. First, the communication architecture is more relying on the open standard communication protocols. The use of the open communication protocols renders the system more vulnerable to cyber attacks in many applications. Second, the SCADA network is moving toward being connected to corporate networks for convenience and other business reasons. Thus the SCADA network may open its doors to outsiders who can enter the corporate networks maliciously.

1.4 Distribution Automation System (Das)

It is a client server technology that provides capabilities for a central server to collect operation data such as voltage and current, to monitor and control feeder remote terminal units which are dispersed in the remote areas, and to detect and restore faults automatically. As information exchange between the DAS server and field equipments becomes more critical for the system operation, communication technology plays an integral part of the distribution system.

1.4.1 Feeder remote terminal unit (FRTU)

They are rather treated as sensors are deployed at remote locations and have the capability of processing data of its own. They also have the capability to detect and restore faults automatically

1.5 Das Communication Architecture

Two integral components of the distribution automation system are the DAS server and the FRTUs. Following figure shows the current fiber-based communication network As shown in this figure, the DAS server and FRTUs are connected to optical ring via modems with a speed of E1 (2 Mbps). A DAS server is connected to a modem through Ethernet while FRTUs are connected through serial ports. The DAS server and each FRTU exchange DNP 3.0 messages on a one-to-one basis. Normally, the DAS server is deployed in a protected area, while FRTUs are placed in untrusted sites as an unmanned system. The communication between the server and FRTUs are not secure since traffic is exposed to the outside of the system, and unwanted traffic can be injected and replayed. Wireless communication is also used in some areas. This kind of communication is basically insecure. Because of its broadcast property, traffic is more vulnerable to malicious access of outsiders. In the current communication architecture, each FRTU cannot exchange information directly each other. Instead the DAS server, acting as a switching hub, delivers data between the FRTUs. But, in order to improve performance. Enhanced services in the distribution system, a decentralized communication architecture will emerge to offer capability that each FRTU can exchange information directly without any intervention of the DAS server. In this communication mode, traffic between the FRTUs will be more vulnerable to various kinds of cyber attacks.

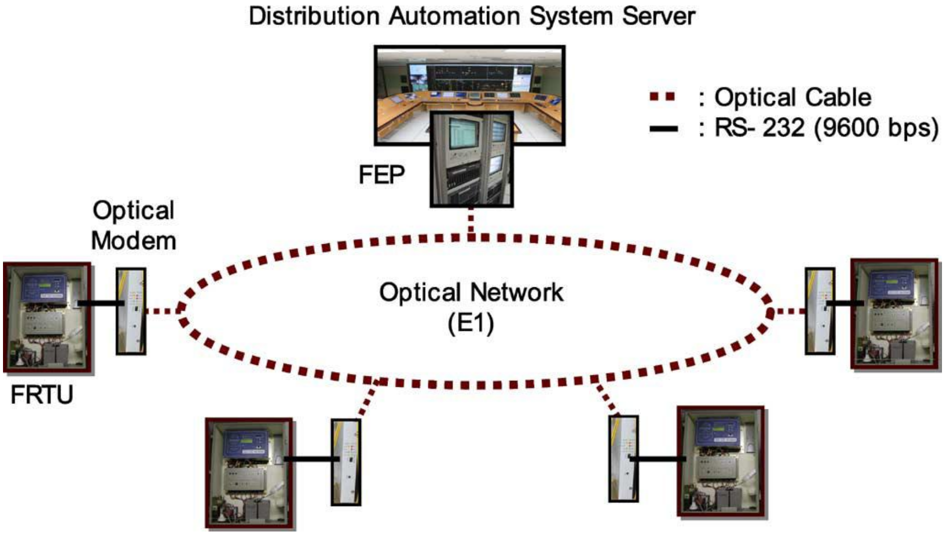

Fig1 DAS Architecture

1.6 Cyber Threats In The Das Network

1.6.1 Denial Of Service Attack

One of the typical network-related attacks to the server is the denial-of-service (DOS) attack. The DOS attack renders the services of the server unusable to the FRTUs. Generally the DOSattack is possible by generating excessive load to the server and consequently exhausting its computing resources. In some cases by taking over legitimate nodes, attackers can swamp the server with unwanted messages. As passive attacks to servers, attackers use malicious codes such as virus and worms to cause malfunctions or halt their functions partially.

Servers can be recovered by rebooting or some other methods when they cannot function properly. Normally these recoveries actions can be taken in a short time since servers are always cared by authorized operators. In this sense, the damage on servers would have little impact on the functions of FRTUs and it is unlikely to cause any severe damage such as power outage to the system. Compared to servers, attacks to FRTUs will make more dangerous effects since they are directly responsible for operations in the field, and are installed mostly unattended in remote sites.

1.6.2 Eavesdropping

The contents of messages that are exchanged between the server and FRTUs can be leaked to outsiders. Eavesdropping is a typical attack of this kind. Attackers can also collect traffic and guess indirectly inside information of the system by analyzing traffic pattern. Messages exchanged between the server and FRTUs contain operating data such as voltage and current, and control commands. Even though information about operation data or commands is exposed to outsiders, this information leakage would not lead critical damage directly to the system operation unless FRTUs are forced to function improperly. In some applications, messages can deliver highly sensitive information such as secret keys which should be known to only the concerned parties. In this case, we need to protect the message contents from eavesdropping. The most dangerous attacks in the distribution system are to cause FRTUs to fail to work properly. There are three distinctive attacks to lead FRTUs to malfunctions.

1.6.3 Content alteration

The first one is to alter the contents of the messages exchanged between the server and FRTUs and then to deliver these false messages to the FTRUs. The modified messages can control automatic switches in the system maliciously, eventually causing power outages.

1.6.4 Creation of bogus message

The second one is to create bogus messages and inject them in the communication channel. Attackers can disguise themselves as the server or they can intercept the communication session. Either way, attackers can deliver illegal commands to the FRTUs.

1.6.5 Replay attack

The third one is the replay attack. All messages contain time varying information which reflects current system status and actions required. Attackers can catch some messages and deliver the messages afterwards. This replay attacks can also make FRTUs to lead malfunctions.

1.7 Requirements For The Das Network Security

1.7.1 Message confidentiality:

Message leakage is not so critical as message modification. In Some applications, however, the message contents should be secured not to be read by illegitimate nodes. The typical application is the secret key distribution where the secret key should be delivered in secure ways. For this purpose, the message should be encrypted by a symmetric or asymmetric key which only intended parties share with.

1.7.2 Message authentication:

Receivers need to verify that messages are sent from claimed senders. Adversaries can inject malicious messages to the FRTUs, consequently causing malfunctions. To authenticate the owner of messages is one of the most important security requirements in many applications in the distribution system.

1.7.3 Message integrity:

Receivers need to make sure that messages they receive are not altered on the way by adversaries.

1.7.4 Message freshness

Freshness means that the message is recent, and old messages are not replayed by any adversary. There are two types of freshness. Weak freshness keeps ordering of the messages but not delay information, while strong freshness not only provides full ordering of the messages but also allow for delay estimation. Weak freshness is required for the application where preventing message replay attack is of main concern, but strong freshness is needed for the application such as time synchronization within network.

1.7.5 Availability:

Services in the distribution network should be always available to all nodes. Servers especially should function properly all the time as they are originally intended. The denial-of-service attack is a typical threat to impair the availability of servers. A single measure cannot solve all security threats. At the same time, the approach which makes all components and their resources be secured is unrealistic since this approach makes the security measures too costly. It is desirable to decide the priorities of what need to be secured taking into consideration the application types and their characteristics in the whole system. In many applications, to hide the message contents by encryption is not so critical. One exception is when the server distributes secret keys to the FRTUs. Message authentication and integrity is far more important than message confidentiality in the applications we consider in the distribution network.

1.8 Message Authentication Code

For message authenticity, a sender generates an authentication tag which is much shorter than the original message and appends it to each message for transmission. A receiver can verify that the message is not altered and the source is authentic by checking the authentication tag. The appended authentication tag is called the message authentication code (MAC). When A sends a message to B, A generates MAC by applying a function to the message and the secrete key between A and B;

The one-way hash function can be used to generate MAC. A hash function, H takes a variable-size message M as input and calculates a fixed-size message digest (MD) as output;

The encryption algorithms can be used to generate MAC from MD. In some applications such as electronic transaction, the asymmetric key algorithm is used to generate MAC, which is known as the digital signature. In this case, if a message is encrypted by A's private key, then B can verify that the A really sent this message by decrypting the message with the A's public key.

The symmetric encryption algorithm is also applied to MD in order to generate the authentication code, where K is a symmetric key. Then A sends the concatenated message of M and MAC.

These approaches have an advantage in that they do not need to encrypt the whole message, consequently reducing the computation cost. However, there is an alternative way to obtaining MAC without involving any encryption algorithm. In this case, A and B share a secret key which is appended to the original message when a hash function is applied;

A sends the concatenated message of M and MAC excluding B computes MAC by applying the same hash function to the concatenated message of M . When B calculates the same MAC as it receives, it is assured that the message must have been sent by the claimed sender A, but also the message has not been altered in transit. The message authentication without encryption is preferable in the application we are focusing on in the DAS networks. The applications we consider in the distribution network are done in the two-way communication between a server and FRTUs based on the multi-access network, either the optical ring or wireless network. Thus, the same message is broadcasted to a number of FRTUs. It is cheaper and more reliable to have each node responsible for message authenticity, and being involved in less computation whenever it receives messages.

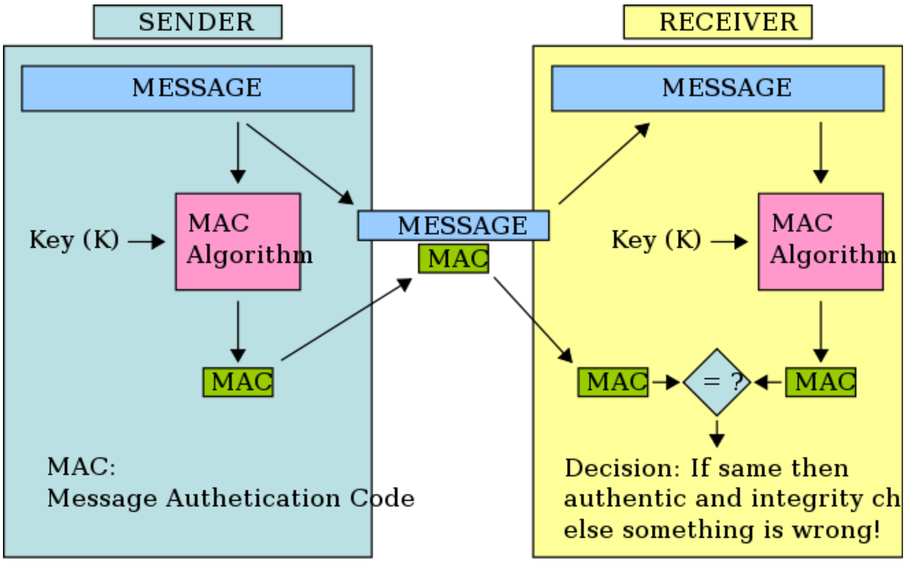

Fig2 Hmac

1.8.1Hashing message authentication code (HMAC)

The message authentication code (MAC) is used to verify the authenticity of the sender and the integrity of the message. In order to avoid computational overhead of any encryption technique, ether symmetric or asymmetric, we choose Keyed-Hashing for Message Authentication (HMAC) as an authentication algorithm. First, the sender A concatenates the Sync Code C, the original message , and the session key, then computesMAC by applying a one-way hash function, H, to the concatenated message. Next, the sender replaces the session key by the MAC and finally delivers the message.

The procedure is as follows:

The node B applies the same hash function to obtain new MAC on the message it received with the shared session key. If these two MACs are the same, B can trust that the message was sent by the claimed sender A, and also the message was not modified on the way. The session key can be of any length. But it is recommended that the key length should not be less than L bytes which is the byte-length of the hash function output, since it would decrease the security strength. Keys longer than L bytes are acceptable but the extra length would not provide significant increase of security .In this protocol we use as a default secrete key length since the default keyed hash function is MD5. The Sync Code is used to verify the weak freshness of the message. Since the Sync Code is a non-decreasing number, the value of a new message should be bigger than the one of an old message. Comparing these two values reveals resent or not, thus ensuring that no attackers replay old messages.

1.8.2The HMAC algorithm in brief

The HMAC algorithm is a shared-key algorithm that uses hash functions for authentication. The strength of the algorithm is based on the strength of the underlying hash function. It uses a secret key and an un-keyed hash function to compute the MAC.

The main operation of the HMAC algorithm is given by :
Let:

- **H**(·) be a cryptographic hash function
- *K* be a secret key padded to the right with extra zeros to the input block size of the hash function, or the hash of the original key if it's longer than that block size
- *m* be the message to be authenticated

- ‖ denote concatenation
- ⊕ denote exclusive or (XOR)
- opad be the outer padding (0x5c5c5c…5c5c, one-block-long hexadecimal constant)
- ipad be the inner padding (0x363636…3636, one-block-long hexadecimal constant)

Then **HMAC**(K,m) is mathematically defined by

$\mathbf{HMAC}(K,m) = \mathbf{H}((K \oplus \text{opad}) \parallel \mathbf{H}((K \oplus \text{ipad}) \parallel \text{m}))$.

```
function hmac (key, message)
    if (length(key) > blocksize) then
        key = hash(key) // keys longer than block size are shortened
    end if
    if (length(key) < block size) then
        key = key || [0x00 * (blocksize - length(key))] // keys shorter than block size are zero-padded('||' is concatenation)
    end if

    o_key_pad = [0x5c * block size] ⊕ key // Where block size is that of the underlying hash function
    i_key_pad = [0x36 * block size] ⊕ key // Where ⊕ is exclusive or (XOR)
    return hash(o_key_pad || hash(i_key_pad || message)) // Where '||' is concatenation
end function
```

II. Existing System

The objective of the present work is

- It is proposed to consider possible cyber attacks in the applications based on the current distribution communication architecture, and then derive the security goals.
- It is proposed to analyze the cryptographic algorithms and devise an efficient security protocol that can be adapted to achieve these security goals, considering the constraints imposed on the distribution system.

The existing system uses symmetric key encryption and HMAC.The HMAC used here has a little of computational over head, old coding techniques and not as effective as that of recent quadruple vector algorithm.

III. Proposed System

The proposed method describes a novel fuzzy class-association-rule mining method based on GNP and its application to intrusion detection. By combining fuzzy set theory with GNP, the proposed method can deal with the mixed database that contains both discrete and continuous attributes. Such mixed database is normal in real-world applications and GNP can extract rules that include both discrete and continuous attributes consistently. The initiative of combining association rule mining with fuzzy set theory has been applied more frequently in recent years.

The original idea comes from dealing with quantitative attributes in a database, where discretization of the quantitative attributes into intervals would lead to under- or overestimate the values that are near the borders. This is called the sharp boundary problem. Fuzzy sets can help us to overcome this problem by allowing different degrees of memberships. Compared with traditional association rules with crisp sets, fuzzy rules provide good linguistic explanation. Also the proposed system uses quadruple vector algorithm.

- **The features of the proposed method are summarized as follows.**

1) GNP-based fuzzy class-association-rule mining can deal with both discrete and continuous attributes in the database, which is practically useful for real network-related databases.
2) Sub attribute utilization considers all discrete and continuous attribute values as information, which contributes to avoid data loss and effective rule mining in GNP.
3) The proposed fitness function contributes to mining more new rules with higher accuracy
4) The proposed framework for intrusion detection can be flexibly applied to both misuse and anomaly detection with specific designed classifiers.
5) Experienced knowledge on intrusion patterns is not required before the training.
6) High detection rates (DRs) are obtained in both misuse detection and anomaly detection.

3.1anomaly Detection

Anomaly detection is a technique of detecting any deviant behavior of the computer whose behavior is being monitored by the system. This paper will discuss two different ways of anomaly detection:

- Profile Based Anomaly Detection
- Signature Based Anomaly Detection

3.1.1Profile Based Anomaly Detection

A behavior is called anomalous if it deviates significantly from the normal behavior. So for anomaly detection, the first thing that needs to be done is to learn the profile of the system. Most precious work done in the area of anomaly detection has used the profiles for user behavior and uses this profile for matching against the system behavior. The anomalies are detected using these profiles which are either statistical in nature or they are learnt using some machine learning technique like neural networks. Another method of building system profile as described in is by building a profile of every root level process running on the system and monitor every process for possible anomalies. This paper describes a way to build profile for any privileged process using the sequence of system calls it make. In general, any privileged process makes the same sequence of system calls at any point in time. This method builds a pattern of system calls and tries to find an anomalous pattern.

Advantages:

- Possible to detect new viruses and worms because this method is not dependent
- Does not need any update or patches on the advent of a new virus.
- Possible to detect viruses and worms which don’t reside on the file system.

Disadvantages:

- Difficult to describe a heuristic which will work on all kind of computer systems.
- Probability of false alarms are high, both for false negative and false positive depending on the threshold.

3.1.2 Signature Based Anomaly Detection

Another method of anomaly detection is signature based anomaly detection which is used by most of the commercial antivirus available today. They have a database of signatures of all the known viruses and worms and they try to match these signatures in any file they find doubtful of infection.

Advantages:

- Probability of false alarms is extremely low.
- Perfect detection of the known viruses and worms.

Disadvantages

- Detection of new viruses or worms is not possible.
- Cannot detect any viral activity which is not present in the file system.

3.2 Data Encryption

It is proposed to apply an innovative technique for data encryption based on the random sequence generation using the recurrence matrices and a quadruple vector. The new algorithm provides data encryption at two levels

Level 1: Transposition

Level 2: Encrypt using recurrence matrices and a quadruple vector.

Hence security against crypto analysis is achieved at relatively low computational overhead.We propose to use a different recurrence matrix and quadruple vector to achieve this proposed objective.

3.3 Quadruple Vector Algorithm

Following quadruple vector Algorithm is used for implementing the process

1. A recurrence matrix used is as a key. Let it be A..
2. Generate a “quadruple vector” T for 44 values, i.e, from 0 to 255.
3. Multiply r= A *T;
4. Consider the values to mod 4.
5. A sequence is generated using the formula [40 41 42]*r.
6. This sequence is used as a key

7 Convert the plain text to equivalent ASCII value.

8 Add the key to the individual numerical values of the message

9 New offset the values using the offset rules
10. This would be the cipher text generated
11 For Decryption the key is subtracted from the cipher text and use the offset rule to get the original message.

3.5 Framework Of Gnp

GNP is one of the evolutionary optimization techniques, which uses directed graph structures instead of strings and trees. The phenotype and genotype expressions of GNP are shown in Fig. 1.2 GNP is composed of three types of nodes: start node, judgment node, and processing node. Judgment nodes, *J*1, *J*2 , . . . , *Jm* (*m* is the total number of judgment functions), serve as decision functions that return judgment results so as to determine the next node. Processing nodes, *P*1, *P*2 , . . . , *Pn* (*n* is the total number of processing functions), serve as action/processing functions. The practical roles of these nodes are predefined and stored in the function library by supervisors. Once GNP is booted up, the execution starts from the start node, then the next node to be executed is determined according to the connection between nodes and a judgment result of the current activated node.

Fig3 also describes the gene of a node in a GNP individual. NT*i* represents the node type such as 0 for start node, 1 for judgment node and 2 for processing node. ID*i* serves as an identification number of a judgment or processing node, for example, NT*i* = 1 and ID*i* = 2 represents node function *J*2 . *Ci*1, *Ci*2 , . . . , denote the node numbers connected from node *i*. The total number of nodes in an individual remains the same during every generation. Three kinds of genetic operators, i.e., selection, mutation, and crossover, are implemented in GNP.
1) Selection: Individuals are selected according to their fitness.
2) Crossover**:** Two new offspring are generated from two parents by exchanging the genetic information. The selected nodes and their connections are swapped each other by crossover rate *Pc* .
3) Mutation**:** One new individual is generated from one original individual by the following operators. Each node branch is selected with the probability Pm1 and reconnected to another node. Each node function is selected with the probability Pm2 and changed to another one.

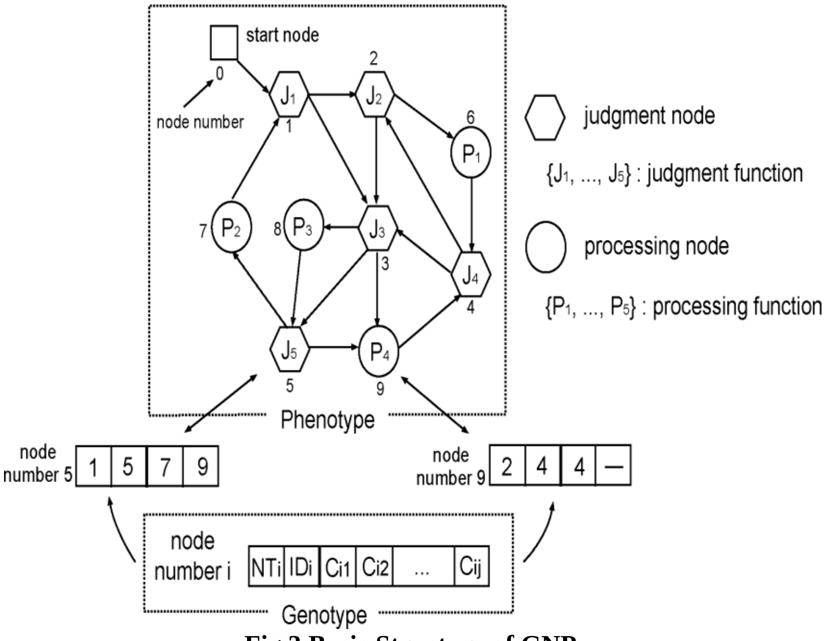

Fig 3 Basic Structure of GNP

3.6 Class – Association Rule Mining

The following is a statement of association-rule mining. Let I = {A1, A2, . . . , Al} be a set of literals, called items or attributes. Let G be a set of tuples, where each tuple T is a set of attributes such that T $\subset$I. Let TID be an ID number associated with each tuple. A tuple T contains X, a set of some attributes in I, if X $\subset$ T. An association rule is an implication of the form X $\Rightarrow$Y, where X $\subset$ I, Y $\subset$I, and X $\cap$ Y =$\emptyset$. X is called antecedent and Y is called consequent of the rule. If the fraction of tuples containing X in G equals x, then we say that support(X) = x. The rule X $\Rightarrow$Y has a measure of its strength called confidence defined by support(X $\cup$ Y)/support(X).

Calculation of χ2 value of rule X ⇒Y is shown as follows. Assume support(X) = x, support(Y) = y, support(X ∪ Y) = z, and the total number of tuples is N. We can calculate χ2 as

$$x^2 = \frac{N(z-xy)^2}{xy(1-x)(1-y)}$$

If χ2 is higher than a cutoff value, we should reject the assumption that X and Y are independent (3.84 at the 95% significance level or 6.64 at the 99% significance level).

Let Ai be an attribute in a database with value 1 or 0, and k be class labels. Then, a class-association rule can be represented by

$$\mathbf{(Ap = 1) \wedge \cdots \wedge (Aq = 1) \Rightarrow (C = k)\ k \in \{0, 1\}}$$

as a special case of the association rule X ⇒Y with fixed consequent C. In this paper, class-association rules satisfying the following are defined as important rules:

□2 > □2 $_{\mathbf{min}}$
support □ sup$_{\mathbf{min}}$
confidence □ conf$_{\mathbf{min}}$

Where χ2 $_{\text{min}}$, sup$_{\text{min}}$, and conf$_{\text{min}}$ are the minimum χ2 , minimum support, and minimum confidence, respectively given in advance.

3.6 Gnp Based Class-Association-Rule Mining

How to Represent Association Rules*:* A judgment node in GNP has a role in checking an attribute value in a tuple. Candidate class association rules are represented by the connections of judgment nodes. An example of the representation is shown in Fig. 4. Processing node *P*1 serves as the beginning of class-association rules. *A*1 = 1, *A*2 = 1, and *A*3 = 1 denote the judgment functions. If a tuple satisfies the condition of the judgment function, Yes-side branch is selected and the condition of the next judgment function is examined in order to find longer rules. No-side is connected to processing node *P*2 to start examining other rules. Therefore, the branch from the judgment node represents the antecedent part of class-association rules, while the fixed consequent part can be predefined.

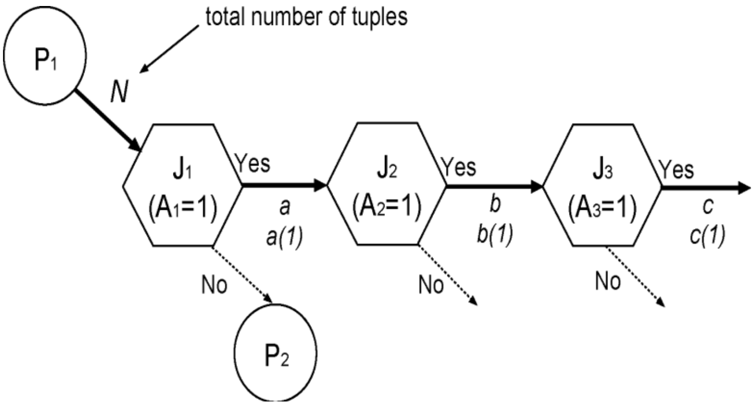

Fig 4 Node transition to find class-association rules.

For example, the class-association rules such as

(*A*1 = 1) ⇒(*C* = 1)
(*A*1 = 1) ∧(*A*2 = 1) ⇒(*C* = 1)
(*A*1 = 1) ∧(*A*2 = 1) ∧(*A*3 = 1) ⇒(*C* = 1)
(*A*1 = 1) ⇒(*C* = 0)
(*A*1 = 1) ∧(*A*2 = 1) ⇒(*C* = 0)
(*A*1 = 1) ∧(*A*2 = 1) ∧(*A*3 = 1) ⇒(*C* = 0)

are examined by the node transition in Fig.4

The procedure of examining tuples is as follows. The first tuple in the database is read and the node transition starts from processing node *P*1 . Then, if Yes-side branch is selected, the current node is transferred to the next judgment node. If No-side branch is selected, the current node is transferred to processing node *P*2 to find other rules. The same procedure is repeated until the node transition started from the last processing node *Pn* is finished. After examining the first tuple in the database, the second tuple is read and the node transition starts from processing node *P*1 again. Finally, all the tuples are examined by repeating the above node transitions. Note that the number of judgment functions (*J*1, *J*2 , . .) equals the number of attributes (*A*1, *A*2 , . ..) in the database.

How to Calculate Measurements of Association Rules: In Fig. (4), *N* is the total number of tuples. *a*, *b*, and *c* are the numbers of tuples moving to Yes-side at judgment nodes *J*1 , *J*2, and *J*3 , respectively. *a*(1), *b*(1), and *c*(1) are the numbers of tuples in class 1 moving to Yes-side at the judgment nodes, respectively. Actually, the processing node from which the node transition starts saves the counted numbers and calculates the measurements. For example, in the case of the rule (*A*1 = 1) ⇒(*C* = 1), the support is *a*(1)/*N* and the confidence is *a*(1)/*a*. In the case of (*A*1 = 1) ∧ (*A*2 = 1) ∧ (*A*3 = 1) ⇒(*C* = 1), the support is *c*(1)/*N* and the confidence is *c*(1)/*c*. *χ*2 values are also calculated by using the above values.

In addition, the extracted rules that satisfy the conditions of the measurements are stored in an association rule pool every generation. Therefore, the rule extraction of GNP is carried out throughout the generations in order to create the rule pool with sufficient rule.

IV. Performance Evaluation

4.1 Misuse Detection

The proposed method for misuse detection is carried out with KDD99Cup database in order to compare with other machine-learning methods. The training dataset contains 3342 connections randomly selected from KDD99Cup database, among which 1705 connections are normal and the other 1637 connections are intrusion, where three types of attacks (nepture, smurf, and portsweep) are included. A total of 41 attributes are included in each connection; however, after the attribute division described in 113 subattributes are assigned to the judgment functions in GNP. After 1000 generations, 3353 rules are extracted.

In the proposed data mining algorithm, each rule is extracted only when it occurs frequently with statistically significant level in the database. Therefore, each rule is not extracted from each connection, but from the whole database by taking account of all the connection data. Therefore, a testing simulation is carried out by the significant rules extracted from the training database. In this paper, the number of connections and the number of extracted rules become similar, but they have no relationship. Actually, in other problems, they are totally different.

The testing database contains 750 unlabeled normal connections and 240 unlabeled intrusion connections (the same types as the training database). The detection results obtained by the proposed misuse detection classifier, where *T* represents the label of the testing results given by the classifier and *C* represents the correct label. Three criteria are used to evaluate our testing results, i.e., DR, PFR, and NFR. DR means the total DR, PFR means the rate at which the normal data are labeled as intrusion, and NFR means the rate at which the intrusion data are labeled as normal.

DR = (746 + 231)/990 = 98.7%
PFR = 4/750 = 0.53%
NFR = 9/240 = 3.75%.

Compared with the results obtained by other machine-learning techniques dealing with KDD99Cup, it is found that the proposed method for misuse detection provides higher DR than most of the machine-learning techniques except the combination method of support vector machine (SVM) with GA and SVM with fuzzy logic. In the aspect of PFR, the proposed method also shows a competitive result. In the future work, the combination of our method with SVM could be considered to improve the performance.

The result of the proposed method is based on one simulation, but actually in the proposed data mining algorithm, the best individual at the last generation is not the solution for the testing, which is a different mechanism from the general evolutionary computation framework. The rule extraction is carried out throughout the generations and all the individuals in a population cooperatively extract many rules. In addition, GNP can only extract rules that are statistically significant. Therefore, the effective rules are stably extracted in every simulation.

4.2 Anomaly Detection

The proposed method for anomaly detection is evaluated by the simulations with DARPA98 database. The training database is intrusionfree for the purpose of the anomaly detection. It contains 9137 normal connection records. After preprocessing, 30 attributes are included in every connection record. However, after the attribute division, 82 subattributes are assigned to the judgment functions in GNP. After 1000 generations, 5589 rules related to the normal connections are extracted.

The number of extracted rules versus generation, which indicates that the proposed method can extract rules of the normal connections efficiently through the generations. The testing database contains 773 connection records including 194 unlabeled normal records and 579 unlabeled intrusion records. Because the training database is intrusion-free, all kinds of intrusions such as back, ipsweep, land, neptune, pod, port sweep, satan, smurf, and teardrop are considered unknown. After the classification using the proposed anomaly detection classifier, the testing results under different settings.

V. Conclusion

It is expected to identify the most critical security goals in the distribution automation system and propose efficient ways of achieving these goals. The message authentication and integrity is far more important than any other security requirements in the distribution system applications. It is also to propose a simple but efficient secret key distribution protocol which uses the symmetric key algorithm. The proposed protocols impose a negligible computation burden on FRTU, resulting in less time overhead on the DAS operation than the one when the encryption algorithms are used. DAS and FRTU have been replace by server and few clients in a wired networking environment and demonstrate all attacks and their elimination.

Reference

[1]. J. Slay and M. Miller, Lessons Learned From the Maroochy Water Breach. Boston, MA: IFIP Springer, 2007, vol. 253, pp. 73–82

[2]. IT Security Advisory Group, SCADASecurity: Advice for CEOs Dept.Commun. Inform. Technol. and the Arts. Canberra, Australia, 2005.

[3]. President’s Information Technology Advisory Committee, Cyber Security:A Crisis of Prioritization Report to the President, Nat. Coord.Office Inform. Technol. Res. and Develop.. Arlington, VA, 2005.

[4]. F. Cleveland, IEC TC57 Security Standards for the Power System’s Information Infrastructure—Beyond Simple Encryption IEC TC57 WG15 Security Standards ver5, Oct. 2005.

[5]. IEC Technical Committee 57, Data and Communications Security, Part1: Communication Network and System Security-Introduction to Security Issues IEC TS 62351-1, May 2007.

[6]. IEC technical committee 57, Data and Communications Security, Part5: Security for IEC 60870-5 and derivatives IEC 62351-5 Second CommitteeDraft, Dec. 2005.

[7]. DNP User Group [Online].

[8]. V. M. Igure, S. A. Laugher, and R. D. Williams, “Security issues in SCADA networks,” Comput. & Secur., vol. 25, pp. 498–506, 2006.

[9]. M. Hentea, “Improving security for SCADA control systems,” Interdisc.J. Inform., Knowl., and Manag., vol. 3, 2008.

[10]. S. C. Patel and Y. Yu, “Analysis of SCADA Security models,” Int. Manag. Rev., vol. 3, no. 2, 2007.

[11]. S. Hong and S.-J. Lee, “Challenges and pespectives in security measures for the SCADA system,” in Proc. 5th Myongji-Tsinghua University Joint Seminar on Prototection & Automation, 2008.

[12]. L. Pietre-Cambacedes and P. Sitbon, “Cryptographic key management for SCADA systems—Issues and perspectives,” in Proc. Int. Conf. Information Security and Assurance, 2008.

[13]. C. Beaver, D. Gallup, W. Neuman, and M. Torgerson, Key Management for SCADA SANDIA, Tech. Rep. SAND2001-3252, 2002.

[14]. R. Dawson, C. Boyd, E. Dawson, and J. M. G. Nieto, “SKMA-A key management architecture for SCADA systems,” in Proc. Australasian Workshops on Grid Computing and E-Research, 2006.

[15]. I. H. Lim, Y. I. Kim, H. T. Lim, M. S. Choi, S. Hong, S. J. Lee, S.I. Lim, S. W. Lee, and B. N. Ha, “Distributed restoration system applying multi-agent in distribution automation system,” in Proc. IEEE

[16]. PES General Meeting, 2008.

[17]. F. Yu, T.-W. Kim, I.-H. Lim, M.-S. Choi, S.-J. Lee, S.-I. Lim, S.-W. Lee, and B.-N. Ha, “An intelligent fault detection and service restoration scheme for ungrounded distribution systems,” J. Elect. Eng. & Technol., vol. 3, no. 3, 2008.

[18]. S. Hong, I. H. Lim, M. S. Choi, S. J. Lee, C. H. Shin, S. W. Lee, and B. N. Ha, “Evolution of communication networks for distribution automation system in Korea,” in Proc. Advanced Power System Automationand Protection, Apr. 2007

[19]. H. Krawcryk, M. Bellare, and R. Canetti, HMAC: Keyed-hashing for message authentication RFC 2104, IETF, Feb. 1997.

[20]. D. Davies and W. Price, Security for Computer Networks. New York: Wiley, 1989.

[21]. F. Baker and R. Atkinson, RIP-2MD5Authentication RFC 2082, IETF, Jan. 1997.

[22]. R. Rivest, The MD5 Message-Digest Algorithms RFC 1321, IETF, Apr. 1992.

[23]. Information Technology—Security Techniques—Key Management— Part 2: Mechanisms Using Symmetric Techniques, ISO/IEC 11770-2, 1996.