ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 3, Special Issue 5, March 2017

Situational Awareness Security for Smart Grids Using Big Data Analysis

Sathya R[1], B.Anbazhagan[2]

Assistant Professor, Ganadipathy Tulsi's Jain Engineering College, Kaniyambadi, Vellore-632012[1,2]

Abstract – Due to the technological improvements in power systems & communication networks, data processing and security issues have become more challenging. The need to protect the data in the power systems before attack is more efficient than analyzing it after it has been affected. Moreover, there is no need to protect the data always. Thus the situational awareness security mechanism based on the big data has been proposed in smart grids.Advanced Persistent Threats (APT), such as Stuxnet, have broken the power system and caused large losses. In fact, there are still many security flaws in the smart grid. Currently, the security of smart grid has attracted a lot of attentions. In this work, the potential security attacks and possible counter-attack measures are analyzed and studied.

Keywords *- Smart Grids, Big data, APT, SVM, Network Security*

I. I. INTRODUCTION

Different kinds of security vulnerabilities were considered for deploying advanced metering infrastructure (AMI). Also, security factors are considered which are related to confidentiality of user privacy and behavior as well as message authentication for sensing and control. Most of the existing security schemes for smart grids focused on the security protection and detection. Security situational awareness is still an unresolved problem in smart grid.

In fact, many threats occur in a very short time and steer by exiting protection and detection components. These threats usually have huge impacts on power system and disturb the normal activities of cities. In addition, network components (e. g. switches, routers) and security components (e. g. IDS, access control systems) of wide-area power systems in the smart grid can
generate security-related big data. In fact, big data are very useful for realizing security situational awareness. Big data are generated in the processes of the security issues and in the operations of generation, transmission, transformation, distribution, consumption and dispatching of the smart grid. Based on the long term monitoring of the smart grid, security related big data can be formed. Then security situational awareness can be realized based on the analysis of the big data.

This work proposes a security situational awareness mechanism for smart grid based on big data analysis. The security facts from e.g., the primary electrical devices, substation buses, network devices, station controllers, control centers and engineering stations, in smart grid can be collected and reported to the security situational awareness center. Then, the data collected over the long term can be stored and analyzed. To get high accuracy for the big data analysis, fuzzy cluster based association method is used to realize the preliminary analysis and game theory as well as reinforcement learning are introduced for security situational awareness.

II. BASIC PRINCIPLE

The proposed network security situational awareness mechanism can be divided into three parts: 1) the extraction of network security situation factors; 2) network situational assessment; and 3) network situational prediction.

Network security situation factors include static configuration information, dynamic operating information and flow information in networks, etc. Static configuration information contains the basic environmental configuration information which includes information on the network topology information, vulnerability information and state information. Dynamic operating information consists of the basic operating information including the threat information obtained from the collection log and the analysis techniques of various protection measures.

To collect the security related big data, the security situational factor data are collected from the communication network of the system in an electric power corporation. The production area processes the operations of electricity generation, consumption, etc. Based on aforementioned data collection scheme, three kinds of situational factors are used to realize the basic situational factor collection, which are network flow, access control operations and devices states. The situational factor data are collected and stored for a long term. Then the data are analyzed based on the proposed mechanism. Security management device reorganizes the abnormal behavior and restores the potential threats in the network.

III. BACKGROUND

Situation Awareness (SA) means that the environmental factors in a certain time and space are cognized and understood, and that future development trends are predicted. Most of the network attacks are generated using the distributed method, which course difficulties for monitoring and controlling the whole network security situation using a simple data fusion mechanism. In addition, in complex network environments, security situational awareness is a complex nonlinear process because of the randomness and

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 3, Special Issue 5, March 2017

uncertainty. The prediction method based on simple statistical data cannot address the above challenges. However, most existing works on security situational awareness focus on the traditional networks. The security situational awareness of smart remains an unresolved problem. The existing security situation awareness works cannot be used in smart grid directly for the following reasons.

First, the smart grid includes wide area heterogeneous networks which based on various special standards, such as IEC 61850, ISO/IEC/IEEE 21451, Wireless HART, ISA100.11a, etc. Second, unlike in normal communication networks, the information modeling and communications of smart grids are combined closely with the complex behaviors and smart decisions of the power system. Finally, the smart grid currently involves more new network models, such as V2G, which enhance the complexity of the smart grid and enlarge the attack surfaces of the smart grid.

Although some schemes have been proposed for the situational awareness for smart grid, these works cannot address above challenges. Although the scheme has significantly enhanced the operator's situational awareness for operating a very large power grid, it was proposed specifically for the power grid control center, and cannot cover highly complex and smart behaviors of the entire grid. Based on above analysis, it is very important to propose a network security situational awareness scheme for smart grid.
There are two methods used for data analysis. They are:

A. *Association Analysis*

Association analysis is a very important issue for obtaining valuable underlying results from a data set. According to different attributes of the data set, association rule mining can be classified as Boolean, classification and quantitative association rules. In the past, rule mining of Boolean associations has attracted a lot of attentions. However, quantitative attributes, such as integer, category and data attributions, are still the most significant association type in applications. For quantitative attribute association, existing mining algorithms of Boolean association rules cannot be applied. Therefore, there are two methods to realize the rule mining of the quantitative association: 1) transforming the quantitative attributes into Boolean attributes; and 2) proposing new association algorithms.

The attribute domain can be divided into several intervals after discretization. Then each interval is mapped as a Boolean attribute. Thus the mining algorithm of quantitative attribute association is converted to the reasonable division problem of the quantitative attribute domain.

B. *Learning Algorithms for Game*

To date, network security situational prediction has usually employed neural networks, time sequence prediction methods, support vector machine (SVM), etc. Most game theory models focus on the equilibrium problem. In fact, the learning model can provide efficient methods for evaluating and optimizing traditional equilibrium concepts. Here the learning model indicates the learning rules of the game player, and checks the interaction among the players. Some learning algorithms have been studied that can be used in the games, such as, Virtual action, reinforcement learning, Experience Weighted Attraction (EWA) learning, etc. However, there are large errors between the predictions and the real practical observation results for the existing learning algorithms for game theory.

Advanced information and communication technologies bring great benefits to the smart grid. However, cyber security threats also extend from the information system to the smart grid. Various attacks can disturb the normal operation of the power system, and thus have serious impacts on the normal productivities and lives of human being. Recently, the information and communication infrastructures of the smart grid have been attacked frequently.

To address the challenges described above, this paper proposes a security situational awareness mechanism for smart grid based on big data analysis. The architecture of the proposed mechanism is shown in Fig. 1. The security facts from e.g., the primary electrical devices, substation buses, network devices, station controllers, control centers, and engineering stations, in smart grid can be collected and reported to the security situational awareness center. Then, the data collected over the long term can be stored and analyzed. To get high accuracy for the big data analysis, fuzzy cluster based association method is used to realize the preliminary analysis, and game theory as well as reinforcement learning are introduced for security situational awareness.

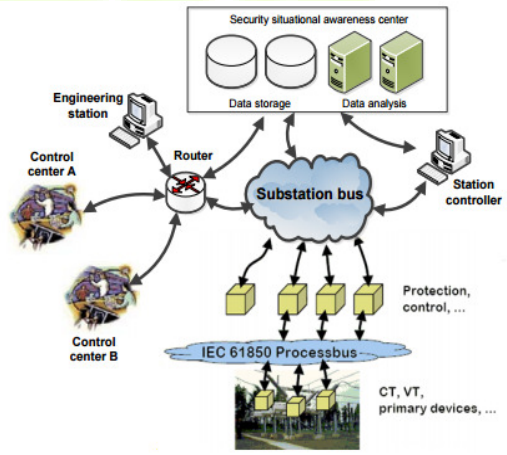

Fig. 1. Security situational awareness architecture for smart grid.

IV. THE PROPOSED SECURITY SITUATIONALAWARENESS MECHANISM

A. *Basic Knowledge*

Situation Awareness (SA), which was clearly proposed by Endsley [7], means that the environmental factors in a certain time and space are cognized and

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 3, Special Issue 5, March 2017

understood, and that future development trends are predicted. The conceptual model of this definition is shown in Fig. 2. However, traditional concept of SA was not originally introduced into the field of network security in the beginning.

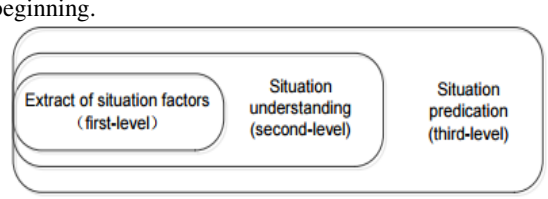

Fig. 2. The concept model of situation awareness.

The network situation can be divided into the security situation, topology situation and transmitting situation, etc. according to application areas. Currently, most existing research works focus on network security situation. M. R. Endsley [7] and T. Bass [8] gave the foundations for network security situation awareness.

Currently, there are still some open issues in the security situation awareness. Most existing schemes do not consider the relevancy of various factors in index architecture, thus making it difficult to fuse all information. Moreover, most of the network attacks are generated using the distributed method, which course difficulties for monitoring and controlling the whole network security situation using a simple data-fusion mechanism. In addition, in complex network environments, security situational awareness is a complex nonlinear process because of the randomness and uncertainty. The prediction method based on simple statistical data cannot address the above challenges. However, most existing works on security situational awareness focus on the traditional networks. The security situational awareness of smart remains an unresolved problem. The existing security situation awareness works cannot be used in smart grid directly for the following reasons. First, the smart grid includes wide-area heterogeneous networks which based on various special standards, such as IEC 61850, ISO/IEC/IEEE 21451, Wireless HART, ISA100.11a, etc [13], [14], [15]. Second, unlike in normal communication networks, the information modeling and communications of smart grids are combined closely with the complex behaviors and smart decisions of the power system. Finally, the smart grid currently involves more new network models, such as V2G, which enhance the complexity of the smart grid and enlarge the attack surfaces of the smart grid. The basic design principle is shown in Fig. 3. Dynamic operating information consists of the basic operating information including the threat information obtained from the collection log and the analysis techniques of various protection measures. The extraction of the proposed network security situation factors has these three advantages: 1) The proposed mechanism can obtain knowledge of all collected information from multi-perspectives; 2) The proposed mechanism considers the relevancy of various factors in an index architecture reducing the difficulties in fusing all information. 3) The proposed mechanism performs effective verification of the index architecture, so we are able to verify the integrated index architecture.

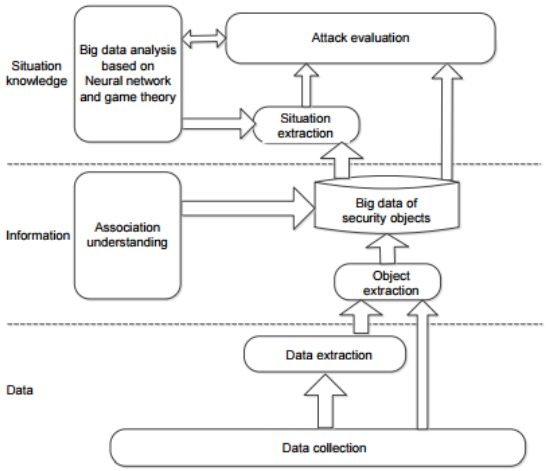

Fig. 3. Design principle of the proposed security situational awareness mechanism.

Understanding the network security situation means fusing mass network security data information and analyzing their relevancy. Based on the acquisition of information described above, the general security situation of the network can be obtained. The process of understanding can also be considered a process of situation evaluation, along with data analysis for network security situation evaluation. In the proposed scheme, network security situation evaluation is not a single security event, and it considers the entire network security state in general, by evaluating the security of the entire network and assisting in decision-making.

To date, network security situational prediction has usually employed neural networks, time-sequence prediction methods, support vector machine (SVM), etc. Most game theory models focus on the equilibrium problem. In fact, the learning model can provide efficient methods for evaluating and optimizing traditional equilibrium concepts. Here the learning model indicates the learning rules of the game player, and checks the interaction among the players.

V. Security Situational Factor Collection and Extraction

The rules of extraction are defined as limitation rules for descriptions of the users' requirements. Situational factors store the security related information and heterogeneous schemes according to the specified format. The inference machine performs the inference operations in the process of requirement parsing, decomposition and optimization. The wrapper and dispatcher perform the packaging and distribution of the task execution. The integrated engine

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 3, Special Issue 5, March 2017

performs the integration of the processing results of different components. In the proposed scheme, semi-structured and unstructured data will be managed uniformly.

In the proposed scheme, three types of situational factors are used to realize the basic situational factor collection: network flow, access control operations and device states. For network flow, we perform the network flowcollection based on Simple Network Management Protocol (SNMP) and the underlying package to capture the flow of the operation systems. Here, SNMP is the standard tool for managing TCP/IP network communications. In the smart grid, SNMP prevails over other commonly used technologies for network management such as common object request broker architecture (CORBA) or network configuration protocol (NETCONF), and has been implemented in most communication architectures. For device state, we also use SNMP technology to realize the situational factor collection. There are three types of components for the device-state situational factor collection including managed devices, agents and Network Management Stations (NMSs) which are deployed in the situational analysis center. A managed device is a node of the smart grid that is used to collect and store the network status, as reported to NMS based on SNMP. Network devices in the smart grid (e. g. routers, servers, switches, network bridges, Hubs, etc.) can act as managed devices. An SNMP agent is a management software component in the managed devices. An SNMP agent collects local information for further management, and transfers the information into compatible format for the SNMP. NMSs are located in the situational monitoring center, which can also provide storage resources for network management. For the access control operational factor, we use the previous method of ours to perform the collection of situational factors. In addition, the features in DARPA 1998 are used as the security situational factors.

VI.EVALUATIONS

A. *Simulations*

To implement the simulations for the proposed security situational awareness mechanism, we use the data set in DARPA Intrusion Detection Evaluation Data [11] for test and training. Eight weeks of network-based attacks of general background data are used for the training. To keep the universality of the data, the midst data of the general background data are used. On the other hand, two weeks attack and background data are used for test. The weights of the proposed awareness mechanism and the constants in the stochastic as well as hidden layers are set according to the corresponding parameters in [12] for comparisons.

The evaluations and comparisons of the awareness are shown in Fig. 4. In Fig. 4, the awareness rate is denoted as the vertical coordinate. Moreover, nine types of attacks in DARPA Intrusion Detection Evaluation Data are denoted as a number of groups of columns, which are IP sweep attack, Mscan attack, Smurf attack, Mailbomb attack, Pod attack, Port sweep attack, Snmpget attack, Back attack, and Teardrop attack. As illustrated in Fig.4, the awareness rate of the proposed mechanism is higher than that of SGA based scheme in [12], and the increment is 9.7% on average, which show the obvious advantages of the proposed mechanism on awareness rate.

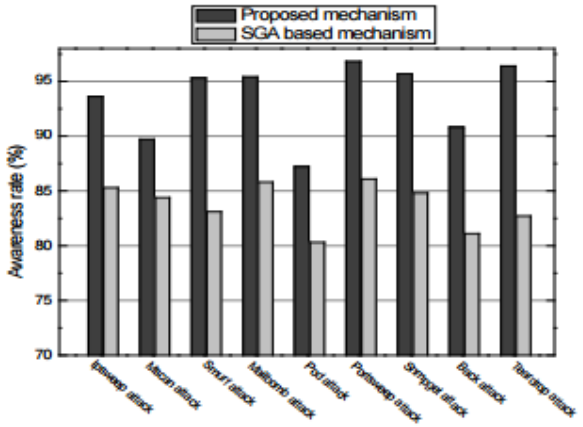

Fig 4. Awareness rate Simulation

B. *Experiment*

To collect the security related big data, the security situational factor data are collected from the communication network of the system in an electric power corporation. The production area processes the operations of electricity generation, consumption, etc. Based on aforementioned data collection scheme, three kinds of situational factors are used to realize the basic situational factor collection, which are network flow, access control operations and devices states. The situational factor data are collected and stored for a long term. Then the data are analyzed based on the proposed mechanism. Security management device reorganizes the abnormal behavior and restores the potential threats in the network.The data analysis experiment is performed on cluster computers with 25 nodes, in which every server has a 2.53 GHz Intel i5 CPU and 8GB of memory. In addition, the servers are connected based on a 1G router. We apply Hadoop-1.0.4 as the experimental tool for the data analysis. The data are collected over half of a year. Here we use the data from the last three months of the collection period (Sept. 16, 2015 to Dec. 14, 2015) to test the awareness mechanism. The risk value RIS is defined to quantify the security situation. Next, we evaluate the error of the security situational awareness. To avoid the negative effects of the large span of the original data, the security factor data are standardized for extrema. For original data Y=(y1, y2, …, yk), the extremum standardization process is computed as follow.

$$y_i' = \frac{y_i - \min(y)}{\max(Y) - \min(Y)}$$

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE)
Vol. 3, Special Issue 5, March 2017

VII. CONCLUSION

Security is one of the key concerns for the smart grid. To realize the security situational awareness mechanism based on the analysis of big data in the smart grid, this paper seamlessly integrated fuzzy cluster based association analysis, game theory and reinforcement learning. Based on the proposed mechanism, the extraction of network security situation factors, networksituational assessment and security situational prediction can be realized for the smart grid. The simulation and experimental results show the high awareness rate and low error rate of the proposed mechanism. The work in this paper is significant for improving the security of the smart grid.

REFERENCES

[1] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart Grid-The New and Improved Power Grid: A Survey," IEEE Communications Surveys & Tutorials, vol. 14, no. 4, pp. 944-980, Forth Quarter 2012.

[2] J. Wu, M. Dong, K. Ota, Z. Zhou, and B. Duan, "Towards FaultTolerant Fine-Grained Data Access Control for Smart Grid," Wireless Personal Communications, vol. 75, no. 3, pp. 1787-1808, Apr. 2014.

[3] X. Wang and P. Yi, "Security Framework for Wireless Communications in Smart Distribution Grid," IEEE Transactions on Smart Grid, vol. 2, no. 4, pp. 809-818, Nov. 2015.

[4] K. C. Sou, H. Sandberg, and K. H. Johansson, "On the Exact Solution to a Smart Grid Cyber-Security Analysis Problem," IEEE Transactions on Smart Grid, vol. 4, no. 2, pp. 856-865, May 2013.

[5] Y. Yan, R. Q. Hu, S. K. Das, H. Sharif, and Y. Qian, "An Efficient Security Protocol for Advanced Metering Infrastructure in Smart Grid," IEEE Network, vol. 27, no. 4, Jul./Aug. 2013.

[6] G. N. Ericsson, "Cyber Security and Power System Communication-Essential Parts of a Smart Grid Infrastructure," IEEE Transactions on Power Delivery, vol. 25, no. 3, pp. 1501-1507, Jun. 2010.

[7] M. R. Endsley, "Design and Evaluation for Situation Awareness Enhancement," Proc. 32nd Human Factors Society Annum Meeting. pp. 97- 101, 1988

[8] T. Bass, A. Arbor, "Multisensor Data Fusion for Next Generation Distributed Intrusion Detection Systems," Proc. of IRIS National Symposium on Sensor and Data Fusion.pp. 24-27, 1999.

[9] S. Jajodia, S. Noel, B. O'Berry, "Topological Analysis of Network Attack Vulnerability,"Proc. of the 2nd ACM symposium on Information, Computer and Communications Security, pp. 2-2, 2007.

[10] R. Xi, S. Jin, X. Yun, and Y. Zhang, "CNSSA: A Comprehensive Network Security Situation Awareness System," Proc. of 2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 482-487, 2011.

[11]DARPA," DARPA 1998 Intrusion Detection,Evaluation,"http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/1998data.html, 1998

[12] T. Kuremoto, M. Obayashi, and K. Kobayashi, "Nonlinear Prediction by Reinforcement Learning," Proc. of International Conference on Intelligent Computing (ICIC 2005), pp. 1-10, 2005.

[13] R. Morello, C. De Capua, "An ISO/IEC/IEEE 21451 Compliant Algorithm for Detecting Sensor Faults: an approach based on repeatability and accuracy, " IEEE Sensors Journal, vol. 15, no. 5, pp. 2541 - 2548, May. 2015.

[14] L. Guo, J. Wu, Z. Xia, and J. Li, "Proposed Security Mechanism for XMPP-Based Communications of ISO/IEC/IEEE 21451 Sensor Networks," IEEE Sensors Journal, vol. 15, no. 5, pp. 2577- 2586, May. 2015.

[15] R. Morello, "Use of TEDS to Improve Performances of Smart Biomedical Sensors and Instrumentation: an overview on advances and applications of ISO/IEC/IEEE 21451 Standard," IEEE Sensors Journal, vol. 15, no. 5, pp. 2497-2504, May 2015.

.