Research of boundary security monitoring model based on Big Data for smart grid

Xiaojuan GUAN[1, a*], Gaofeng HE[1, b], Cheng ZHOU[1, c], Weiwei LI[1, d]

[1]Information and Communication Department, State Grid Smart Grid Research Institute, Nanjing, 210003, China

[a]email:guanxiaojuan@sgri.sgcc.com.cn, [b]email:hegaofeng@sgri.sgcc.com.cn,

[c]email:zhoucheng@sgri.sgcc.com.cn, [d]email:liweiwei@sgri.sgcc.com.cn

Keywords: Smart Grid; Border Security; Network Security; Situation Monitoring.

Abstract. To enhance the defense capability of information security in smart grid and protect the data exchange between intranet and external information network, a kind of safety device has been produced that achieved effective control of information network data access. On the basis of the company 's multi-channel defense system construction , the defense methods of border security were studied and the monitoring model of mass interactive data between intranet and external information network was established to achieve real-time interactive data border monitoring, network behavior collection and retrospective database access, and to provide a basis for the evaluation and prediction of border security status, and to enhance the overall monitoring level of the network information inside and outside boundary state.

Introduction

Smart grid is the inevitable trend of the development of the world power grid and a common trend of the development. As the high integration between information technology and the transmission and distribution infrastructure, the method of the protection and monitoring all aspects of grid operation could be used to enhance the intelligence level of the grid and to implement the goals of the safety, reliable and economic grid.

Since blackout events took place in 2003, the United States has begun to study smart grid related technologies and used new technology to Reform the traditional power network infrastructure and facilities. The European commission released “A European Strategy for Sustainable，Competitive and Secure Energy” [1] in 2006 which clearly emphasized on “Europe has entered a new energy era, smart grid are the future direction of the European power grid”. In 2009, China put forward “strong smart grid” to promote the work from planning, constructing and improving these stages.

With the development of information communication and control technology, smart grid has the characteristics of information, automation and interaction which also brought lots of information security risks. With the popularization and application of smart grid interactive service, data interaction inside and outside the network is getting more and more frequent, so how to ensure the boundary security inside and outside the network is the key problem. A kind of safety device has been produced that achieved effective control of information network data access, but the audit network behavior monitoring has not been achieved.

Massive border interaction information security monitoring model

The current methods of internal and external network border security protection focus on isolation protection, which could detect and trace malicious attacks. In order to adapt to the interconnection trend between the outside network and the Internet widely interconnected trend, on the premise of maintaining border barricade deployment, border security key methods should be researched to achieve real-time interactive data border monitoring, network behavior collection and retrospective database access, which help network administrator to understand security status from point to plane and identify the border security events and security threats and form a border security monitoring system[2][3]. The system is mainly divided into the following several parts:

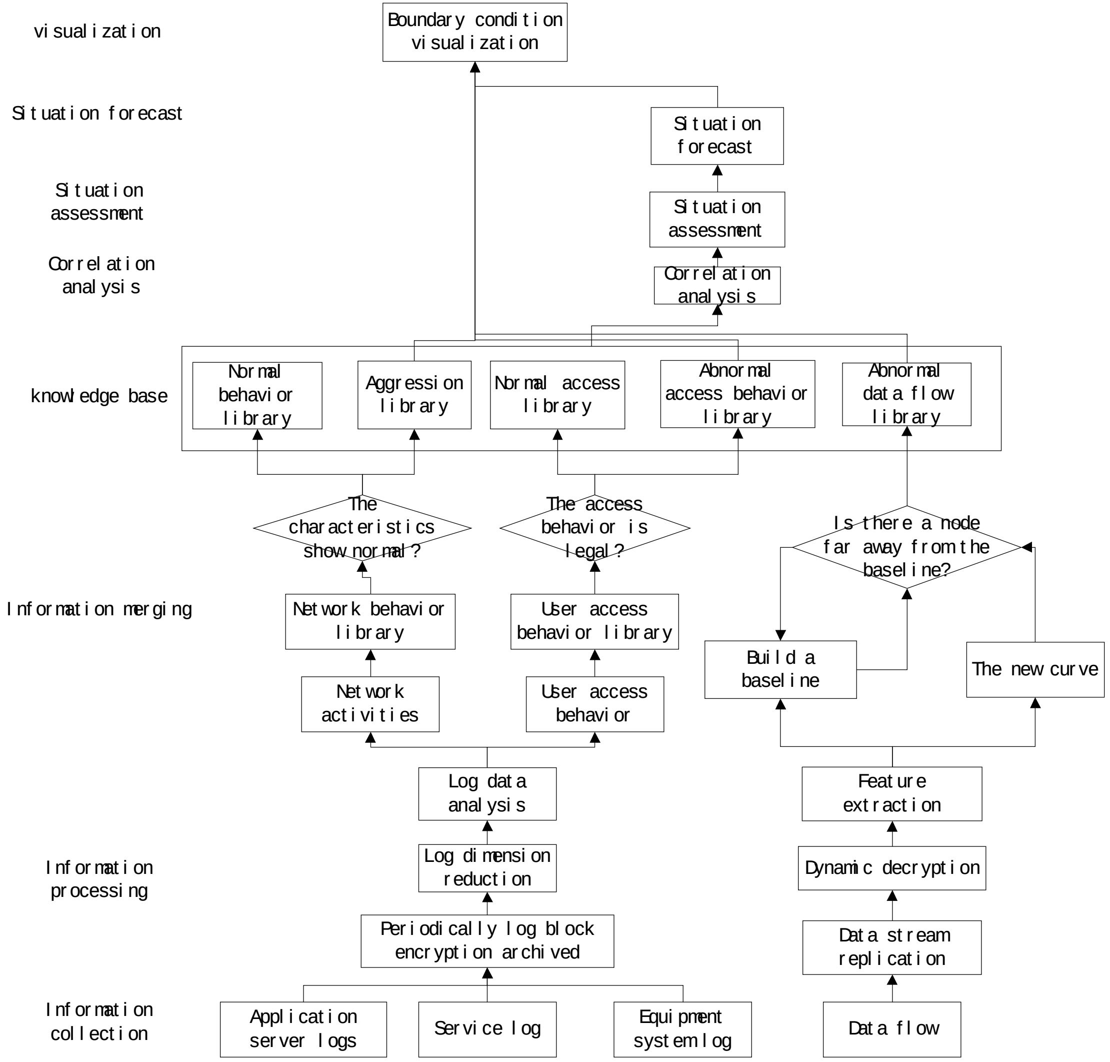

Fig. 1 the boundary security monitoring model

Information collection

The elements include: the logs of device node and the data flow:

- Real-time acquisition of data flow

Data flow is continuous and massive and these data elements can only be read once according to a fixed order, so data stream replication technology can be used to capture specified packets.

The safety device we produced using private transfer protocol and custom encryption algorithm, so after getting data packets, the packets should be decrypted dynamically and filter out the agreement with database stream packets and extract data from the database protocol flow.

- log information acquisition

The application log, system log and service log would be got from multiple servers and the product researched by our company.

Information preprocessing

The object of network behavior includes network behavior, user access behavior, data flow, etc.

Information processing is processing systems log, application log, service log and data flow of network equipment deployed in the network border to implementation dimension and data analysis.

- Log processing

Massive log has been produced in the process of information interaction between inner and outer network in smart grid which had the problem of duplicate redundancy, characteristic dimension and the logic relation cannot be described. The space dimensions and complexity are very high in log events and the growth of log file is the speed of GB / day which make log storage and log analysis difficult. The log information which has been classified simply still can’t meet the needs of users targeted

Log block encryption archive: the storage frequency has been set according to the collected data. The log has been encrypted and stored using the algorithm of asymmetric encryption. Only administrators have the decryption method and the permissions of calling and reviewing the archive logs. An attacker who must obtain administrator privileges can achieve access to the log files.

Log dimension: The duplicate redundant data log data has been merged into one record by analyzing log characteristic value such as the time stamp, source IP, port, transmission content and so on and has been reduced the size and complexity of the log data by order of magnitude of diminishing.

- Data Stream Processing

The traffic data packets on the network have been obtained using stream replication technology to extract the monitoring object, such as the total number of connections, the flow, the number of packets and so on.

Information feature extraction

The log includes the log text and data flow information, now the two types of information are extracted respectively:

- Log feature extraction

Analyze the Transmission content, extract data form the specific request, such as log in, log off, SQL request, large data transmission request and so on. Lexical, syntactic and semantic analysis have been used to form unified network behavior format. All kinds of behaviors have been analyzed to build network behavior library and user behavior library. Data mining and machine learning have been used to process the behavior and build typical behavior activities.

- Build multi-object contour lines

According to the different objects of data flow, the normal curve contour lines have been built. The data packets have been analyzed in real-time using the method of adaptive adjustment parameters, comparing with prebuilt contour lines, abnormal conditions has been detected and recorded into abnormal traffic library. If the actual values are in the normal range, the values will be added into the baseline using the mean algorithm to form a new baseline[4].

Establish the knowledge base

The attacks between the internal and external network include SQL injection, DDOS, virus and Trojan which have their own characteristics. The network data flow has been analyzed to identify abnormal data flow, according to the known characteristics of aggressive behavior, the attack behavior characteristic library has been built to implement statistics and accurate recognition of flow type and the network behavior.

The current common network attacks, such as port scanning, DDOS, flood attack, ping of death, land attack and so on, have their own uniqueness. For example, the feature of port scanning is that the access from the same source address to different destination address or port; the feature of flood attack is that a large number of abnormal packets being sent to the server ; the feature of ping of death is that the bytes of ICMP ECHO exceed 64KB; the feature of land attack is that they have the same source address and destination addresses.

According to the characters of known attacks, by analyzing the network data flow, identifying abnormal data flows, the attack behavior characteristic library has been built to to achieve statistics and accurate identification of the type of traffic and network behavior.

Building knowledge base has been divided into the following steps:

- Normal behavior characteristic library and attack behavior characteristic library have been built.

According to the characteristics of network behavior, the network behavior has been stored into the two libraries. The method based on the rules and machine learning has been adopted to establish the characteristic library containing all network behavior and user access, and the way combined black list with white list has been adopted to determine the nature of network behavior.

- Build normal database access behavior library and abnormal database access behavior library.
- The data streams which have anomaly characteristics have been recorded into the abnormal data flow characteristic library

Correlation Analysis

Correlation analysis is a common method of data mining in the area of intrusion detection, which is used to mine the correlation between data. The main mining algorithms contain Apriori, DHP, AprioriTid, FARM, Auto-Apriori, AVM, StepLength, STBAR, FP-growth and HCS-Mine, and most of them are derived on the basis of Apriori. According to the preset minimum support and confidence threshold, the candidate sets greater than the support and credibility have been dug out.

According to the established knowledge base, multidimensional analysis[5] has been adopted to process relevant information. According to the time stamp value and behavior characteristic value, the method of correlation analysis based on the partition has been adopted to all libraries we mentioned. Based on the concept of A B + A C = A B C, the new network attack sequence has formed.

Information display

Information Visualization makes network status, network environment, access behavior visual. Monitoring content can be realized to classify and display, such as the alarm information, aggressive behavior, the user operation behavior, data flow state information and so on. At the same time, the assess value and the predictive value of the border security state could be display which helps administrators to effectively guide the deployment of security devices inside and outside network.

Conclusion

For the security requirements of the information inside and outside the network border, information security audit model and Data mining framework have been built to form many knowledge bases of characteristics and provide supports for situation assessment, prediction and visualization in next step. Active defense system has been created to enhance the monitoring and auditing level of information inside and outside the network boundary and to facilitate the construction of strong smart grid.

Acknowledgement

In this paper, the research was sponsored by the Basic and Prospective Project of SGCC (SGRIXTKJ[2015]216).

References

[1] EUROPEAN COMMISSION. Energy 2020 A strategy for competitive, sustainable, and secure energy. Publications Office of the European Union,2011.

[2] STATE GRID Corporation of China , CHINA ELECTRNIC POWER RESEARCH INSTITUTE, TIANJIN ELECTRIC POWER CORPORATION. A method of false positive adaptive network security situation prediction: CHINA,201410705040.6[P].2014-11-26.

[3] STATE GRID Corporation of China , CHINA ELECTRNIC POWER RESEARCH INSTITUTE. A method of parameter adaptive network security situation of quantitative evaluation : CHINA , 201410535005.4[P].2014-10-13.

[4] Bluedon Information Security Technologies CORPORATION. A network boundary abnormal monitoring method:CHINA, 200910214525.4[P]. 2011-07-06.

[5] Wu Guo-qiang. Situation of Technical Analysis and Evaluation of Network Security Event Correlation[J]. Information Security and Technology.2014(12):15-16,47.