Design of Flow Analysis System Based on Industrial Control Network in Tobacco Industry

Chengting Zhang, Jie Qian, Xun Zhang and Qi Xu*
China Tobacco Zhejiang Industrial Co. Ltd, Ningbo 310009, China

*Corresponding author email: qjztc@163.com

Abstract. This paper has introduced a system dealing with network flow in tobacco industrial control system. By effective data acquisition, data processing, information analyzing and data storing, the system provides precise information for network security analysis. In this article, each key link of the processing chain such as acquisition, processing, analyzing, and storing has been specified. Besides, the feasibility and practicability of this system has been proved. The system would help ensure the safety and stable operation of tobacco industrial control system.

Keywords: Tobacco industrial control system, Flow analysis, Data acquisition, Network security.

1. Introduction
Nowadays with the deepening integration of information and industrialization, information technology is increasingly used in tobacco industrial control systems. The core systems, represented by the MES, have been connected due to the demand of management and data interaction. The industrial control network becomes a vital part of the cigarette factory network, connecting internal and external networks. The rise and development of the Internet of Things has brought a boom in the close combination of industrial control systems and the Internet, which has greatly promoted the development of industrial. However, at the same time, it inevitably brings about information security risks. Firstly, with the deepening integration of information and industrialization, industrial network has been interconnected with office network and internet, the security of some core systems in industrial control system is directly related to the national information security. Secondly, core business mostly deployed in the industrial control network, which is highly informationized and irreplaceable. The influence of operation security problem is obvious. Thirdly, the legal norms and national security standards in the field of industrial control system are relatively lacking, and there is no strict market access system, in addition, the state's industrial support for domestic industrial control equipment needs to be strengthened. The situation must be completely changed. Otherwise, national security and the well-being of the people will be shrouded in the shadow of the security risks of the industrial control system.

2. System Objectives
In order to solve the shortages of industrial control safety protection in tobacco industry, the system is designed to collect network data and monitor security threats by using data acquisition and analysis technologies. The system is to achieve the goals that realize the ability of format storage and visual analysis of traffic key communication data of industrial control network.
3. System Design
Tobacco industrial network analysis system which monitors network security involved five steps: data acquisition, data preprocessing, data analysis, data storage and data return.

3.1. Data Acquisition
The data acquisition function provides the basic data source for the network behavior analysis, including data-packet analysis and reorganization, so that it can detect abnormal behavior through the 7-layer protocol of industrial control network. Data acquisition realizes the detailed analysis ability of all communication data packets in the network, and realizes multi-level network communication data analysis.

The system introduces sniffing to achieve data acquisition. Sniffing is a method to capture data packet through mirror network flow from core switch or distribute switch. It captures more comprehensive information than SNMP or NetFlow by completely copying data packets. The captured data packets from different protocols such as TCP/UDP/S7/Modbus/Profinet will be stored losslessly.

After the network traffic is obtained by sniffing, the session information of all data streams in the network can be saved through real-time processing of data streams, including physical endpoint session, IP host session, TCP session and UDP session information, application and protocol session information.

Through the storage of network session, users can view and understand the network session information at any time, find abnormal communication session in time, and quickly detect all kinds of network problems.

3.2. Data Preprocessing
The data preprocessing engine completes the preprocessing of the real-time data stream directly obtained from the network. First, the real-time data stream is restored by session in transport layer, eliminate the interference of disorder, retransmission and delay caused by network conditions in order for analysis.

Then the application protocol is identified to determine the specific application carried on the data stream. Finally, the structured metadata information is extracted from the unstructured data stream for subsequent statistical and correlation analysis.

Metadata is the description data of the original message. It does not contain the specific content of the original message but it conserved most of the information in the original data stream which can be used for subsequent statistical analysis. Typical metadata includes a session's start time, end time, source address, destination address, source port, destination port, protocol type, application type, uplink traffic, downlink traffic, packet distribution, etc. When extracting metadata based on 11 tuples, sufficient information should be saved for subsequent analysis module use, and too complicated calculation should be avoided, which reduces the processing efficiency of data flow. The size of data processed after metadata extraction will be much smaller than the original data packet, and the bandwidth of returned data is about 0.9% - 1.5% of the original traffic.

3.3. Data Analysis
When it is found that there is a sudden flow or abnormal flow in the industrial control network, the data analysis can be used to analyze the flow in this period, find the causes of network abnormality in time and avoid further problems. At the same time, for the occurrence of historical problems, it can quickly extract the data of this period for detailed analysis of historical data. Some network problems may not be represented by abnormal traffic, such as the slow response of the database server at a certain time in the past. To analyze the cause of this problem, we need to analyze the communication data of this period. With the capability of long-term data storage, the system can mine and retrieve the historical data of any period in the past and quickly retrieve the historical information and conduct fine secondary analysis to find the cause of the problem.
Figure 1. Data collecting hierarchy.
Collection and analysis provides data mining from multiple perspectives of network protocol and application, physical endpoint, IP endpoint, physical session, IP session, TCP session, UDP session, etc. It can intuitively reveal the relationship between the network objects and the statistical results of data. For example, through a certain protocol, the IP endpoint can be mined, and the session related to the IP. The system also supports skip between different layers, which can be easily traced back to any mining path node.

The analysis model includes four types of network abnormal behavior detection, as follows:

1. Suspicious network behavior detection
   Through the analysis of various traffic parameters of all hosts in the network, the network behavior anomaly detection rule base of hosts is defined to realize the detection and discovery of suspicious network behavior of hosts.

2. Key objects monitoring
   Key Objects is monitored by analyzing the traffic parameters of the network application traffic and the network host traffic in real time, defining the monitoring rules. Key objects monitoring realizes the network communication abnormal alarm of the application traffic and the host traffic.

3. Sudden flow
   It can detect the burst traffic of the link, find the abnormal traffic and alarm.

4. Feature Detection
   Feature data detection is a function of monitoring feature behavior and abnormal communication access in the process of full flow collection by defining the feature matching conditions in the data package.

3.4. Data Storage
The system supports the long-term capture and storage function of high-performance network traffic. When the data reaches the upper limit of local storage capacity, it can be automatically replaced in chronological order to ensure the consistency and real-time of stored data. At the same time, in order to
facilitate the subsequent query and analysis efficiency, the following two types of statistical data are stored in real time:

1. **Network session storage.**
   Through the real-time processing of the data stream, the system can save all the session information of the data stream in the network for a long time, including physical endpoint session, IP host session, TCP session and UDP session information, application and protocol session information.

2. **Link flow parameter storage**
   The system can analyze, count and save the total traffic parameters on the network link in real time, such as total traffic, broadcast/multicast traffic, uplink/downlink traffic, packets, utilization, number of TCP synchronization packets, number of TCP synchronization confirmation packets, number of TCP synchronization reset packets, to help users quickly understand and know the network operation status.

3.5. **Data Return**
To ensure that the flow data after preprocessed can be accurately and timely transmitted to the third-party management system. Under the condition of correct queue for data transmission, the system also designs the synchronous and asynchronous mode of data return.

![Data return](image)

**Figure 3.** Data return.

The system can regularly detect the connection of the interface with the external network. If the connection is normal, the system will return the data through the network. The mode of data return can be adjusted according to the actual situation. Under normal network conditions, users can also export data locally.

If the network is disconnected, the system can store data to local storage immediately. At the same time the system keeps detecting network conditions until the connection recovers. The system allow users to export data locally during abnormal network conditions.

4. **Conclusions**
By implementing this system, it can establish a five layer processing architecture for network traffic collection, preprocessing, analysis, storage and return with clarifying the role of each layer. By processing and storing data effectively, the ability of network security protection significantly improved and realized the security monitoring ability of prejudgment, in-process detection and after diagnosis.
References

[1] Ministry of industry and information technology: Notice on strengthening information security management of industrial control system. Ministry of industry and information technology (451), (2011).

[2] National Tobacco Board: Network interconnection security specification for production and management network of tobacco industry enterprises. National Tobacco Board, (2013).

[3] Shuai Chen: Analysis of computer information processing technology in the background of big data. China New Telecommunications (6), 22-24 (2018).

[4] Rong Lin: Research on Key Technologies of network traffic control. Science & Technology Information (12), 16-18 (2010).

[5] Yongxin Cheng: Development of computer network security protection system [J]. Study on the transmission power, 2020, 4(1):182+184.

[6] Zhaofeng Yang, Aiwan Fan, Tongqian Peng. Research on the construction of computer network security system based on big data environment [J]. Information Technology And Informatization, 2019 (11): 148-150.