Industrial Control Network Anomaly Data Identification Method based on Wireless Communication

Xiaojin Mo

Wuhan Institute of Shipbuilding Technology, Wuhan 430050, Hubei Province

kristymo@163.com

Keyword: Network monitoring; Anomaly data identification; Wireless communication

Abstract: With the gradual igniting of industrial infinite communication technology and the gradual internationalization of development, it plays a pivotal role in the road of industrial automation. Industrial network anomalies have also become a topic of concern, and research on the identification of anomalous data in industrial networks is extremely important. However, most of the data that can be processed by the current industrial network data anomaly identification method is a single dimension, and data detection can only be performed on a single item to be measured, and the method of measuring the network data is huge. This method is extremely inconvenient. And can't measure more rigorously. In order to solve such problems, this paper finds a large number of documents, and through the comparative analysis of data, finds an abnormal recognition method that can detect data of multiple indicators to be tested, and improves the reliability and security of network data identification.

1. Introduction

As the development of network technology continues to grow, the possibility of network anomalies is increasing. Network exception handling is generally used in areas such as network security and network resource protection. Facing such a huge data resource of the network, it is extremely important to improve the security of data and maintain the normal operation of the network. Especially in the intelligent application of the network gradually applied to the industry, to promote the development of industry, the research on the abnormal data identification of industrial networks is urgent.

In recent years, with the continuous research and research by researchers, research on network data anomaly recognition in terms of unlimited communication has been continuously developed. From 2017 Yi Shengwei, Zhang Yibin, Xie Feng [1], etc., in order to study the application of network security, using fuzzy test technology, based on Peach, the security analysis method of industrial control network protocol was proposed, and monitoring was found to be abnormal during the detection process. Analysis, the results of the analysis using the network security protocol confirmed its security. In 2017, Zhang Kaiyi, Chen Tieming, and Yan Chun [2] used the security detection of industrial control systems, the structure, vulnerability, and threat of industrial control systems in order to solve the threats of viruses and Trojans brought about by the development of Internet convergence. Four aspects of anomaly detection were studied, and suggestions on industrial network security protection were put forward. In 2017, Han Dantao, Zhao Yanling, Yan Xiaofeng [3] designed a special industrial network security isolator based on PROFINET to ensure the security of industrial control network, based on industrial communication network diagnosis, isolation and security protection technology, to achieve real-time monitoring of network status and PROFINET key data, blocking abnormal malformed messages, and preventing unauthorized access of unauthorized devices. The alarm information generated by the above situation will be sent to the configuration management platform in real time and alarm display. In 2018, Liu Wanjun, Qin Jizhen, Qu Haicheng [4] used the detection method of industrial control network intrusion detection method to solve the problem of internal abnormal point and outlier point in order to solve the problem of single class support vector machine intrusion detection, through DBSCAN algorithm and K-means

DOI: 10.25236/iciss.2019.004

method. And OCSVM anomaly intrusion algorithm, designed to use the combined classifier to detect industrial network intrusion without abnormal data samples, improve the detection effect. In 2018, Xu Guozhong [5] in order to solve the network security risks caused by the lagging operating environment and difficult implementation of industrial systems, using network data traffic anomaly monitoring and deploying virtual gateways to generate alarms, the status of network information security in industrial control systems Analyze and propose a safe and reliable network operation reinforcement scheme. In 2018, Zhu Jianjun, An Panfeng, Wan Ming [6] in order to solve the complexity of industrial control network environment and the speciality of intrusion detection requirements, using deep analytical anomaly behavior, using rough set theory to simplify detection features, using support vector machine Algorithm classification, combined with adaptive genetic algorithm for model optimization, results in an intrusion detection method that can reduce the complexity and detection time of the intrusion detection model and improve the detection rate. In 2018, Zhao Cheng, Fang Jianhui, Yao Minghai [7] proposed a multi-feature space weighted combined SVM classification detection algorithm to detect APT attack anomalous session flow in order to solve the network security problem of advanced persistent threats, using deep stream detection technology. The detection accuracy is higher, the false alarm rate is lower, and the safety efficiency is improved. In addition, Su Chunlei and others have also conducted research on this [8-10].

Industrial control research based on wireless communication has also been gradually expanded. In 2017, Han Cunwu, Chang Shurui, Qi Qi et al. [11] studied wireless and wireless network re-modeling, control time-delay power and rate control models for wireless communication networks. A method of optimal tracking of communication network power and rate. In 2018, Lu Wei [12] proposed a reliability confidence interval prediction algorithm for wireless communication links based on wavelet neural network and fuzzy control theory in order to study the nonlinear prediction and non-stationary random characteristics of wireless links. The link reliability control algorithm improves the stability and reliability of the link quality. In 2018, Xue Xue, Wang Jianping, Sun Wei [13] in order to explore a method for quality control of cross-layer protocol of wireless sensor network protocol stack in micro-grid data transmission, mathematical modeling, based on the concept of fuzzy cognitive map The quality has made a fuzzy control method, which enables the micro-grid data communication of the wireless sensor network to provide effective QoS guarantee. In 2018, Zheng Dongliang, Li Shichao, Zhang Liting and others [14] in order to study the application of remote monitoring in the industrial field, based on WIA-PA industrial wireless technology development platform, real-time detection of oil well data, greatly improving the production efficiency of oil fields and Management level. In addition, Zheng Dongliang, Zhang Liting, Li Shichao and others have also conducted in-depth research on industrial wireless communication [15-19].

In order to study the anomaly data identification method of industrial control network based on wireless communication, this paper studies the abnormal data identification from three methods: anomaly detection method based on probabilistic PCA model, anomaly detection method based on VB inference and anomaly detection method based on wavelet transform. [20-22]. Through comparative analysis, it is concluded that the PPCA model has a higher detection rate for abnormal data and a lower false positive rate, which greatly improves the reliability and stability of network operation [23-25].

2. Method

2.1. Anomaly Detection Method based on Probabilistic PCA Model.

Probabilistic PCA models are used to improve the performance of PCA processing random data sources.

Let the number of flow OD pairs be P, and the number of flow sampling points in the detection time window be n, then the observation model of the network flow signal is

$$D_{p\times n} = M_{(r)p\times n} + E_{p\times n} \tag{1}$$

Where: $M_{(r)p\times n}$ is the observed signal matrix of the flow, r is the rank, the magnitude of the flow must be normalized and zero-centered and then input into the matrix; $E_{p\times n}$ is the observed noise matrix caused by the random burst, in order to satisfy the computational complexity The demand for real-time detection uses a Gaussian distribution with a mean of 0 and a variance of ω^{-1} to approximate random burst traffic. Therefore, the PPCA model of the network traffic signal is

$$f(D|A, L, X, \omega, r) = N(ALX', \omega^{-1} I_p \otimes I_n) \tag{2}$$

Where: ALX' is the economic singular value decomposition of $M_{(r)p\times n}$, $A \in R^{p\times r}$ and $X \in R^{n\times r}$ are both the western matrix; I_{p} and I_n are the identity matrix; L is the diagonal matrix composed of non-zero singular values, is

$$L = diag(l); l = [l_1, ..., l_r]^T;$$
$$l_1 > l_2 > ... > l_r > 0 \tag{3}$$

2.2. Anomaly Detection Method Using VB Inference.

The rank of the PPCA model reflects the total amount of principal components. The attack traffic on the network has a tendency to increase the rank, while the random burst traffic tends to make the rank distribution tend to be smooth. The basis of this conclusion lies in the following two points: 1. The total principal component of the flow consists of several normal principal components with periodic stability and several abnormal principal components that suddenly appear. The number of abnormal principal components will be caused by the occurrence of attack flow. The increase, while the normal principal component has not changed, so the total number of principal components will increase, and because the rank reflects the total amount of principal components, the attack flow has a tendency to increase the rank; 2. According to formula (2) The random burst traffic can be regarded as the sum of the Gaussian part and the non-Gaussian part. The Gaussian part does not affect the rank distribution because it has been included in the PPCA model, and the non-Gaussian part has the influence on the rank distribution because there is no prior information. This will have a flattening effect on the rank distribution function. Based on the above two reasons, random burst traffic smoothes the rank distribution. To design a detection algorithm based on this principle, it is necessary to infer the rank of the PPCA model. In the Bayesian inference system, the EM and MAP methods based on the maximum likelihood principle use the rank of the model as the a priori information when solving the inference problem of the PPCA model, so it does not have the ability to infer the rank; it has developed in recent years. The variational Bayesian (VB) theory, although computationally large, not only can infer the rank of the model, but also can obtain the distribution function of the rank. Therefore, this paper uses the VB algorithm as the inference algorithm for the rank of the PPCA model.

According to the Bayesian formula, the posterior distribution of r can be expressed as

$$f(r | D) \propto f(D | r) f(r) \tag{4}$$

Approximate by the knot inequality, there is

$$\begin{aligned} &\ln f(D | r) \approx \ln f(D | r) - KL(\tilde{f}(A, X, L, \omega | D, r) \\ &\| f(A, X, L, \omega | D, r)) = \\ &\int_\Theta \tilde{f}(A, X, L, \omega | D, r)(\ln f(D, A, X, L, \omega | r)) \\ &- \ln(\tilde{f}(A, X, L, \omega | D, r)) dAdXdLd\omega \end{aligned} \tag{5}$$

Where: KL(·|·) is the KullBack-Leibler(KL) deviation of two distributions; $\tilde{f}$ is the optimal approximation of the distribution f.

2.3. Anomaly Detection Method based on Wavelet Transform.

The wavelet transform method can provide a time-frequency window that changes with frequency, so that the frequency is automatically widened at low frequencies and automatically narrowed at high frequencies. Because the wavelet transform can convert the data signal to different frequency bands, the slow change of the data signal can be reflected by the low frequency, and the instantaneous change of the data signal can be reflected by the high frequency. After relevant research, ian superimposes a small interference signal on a common waveform. After wavelet transform, the signal at the interference will be abrupt, and the amplitude of the signal will be much higher than normal. In this way, the wavelet transform analysis technique can be used to find related mutation points (singular points). Therefore, in the field of traffic anomaly detection, using the resolution characteristics of wavelet variation, the nonlinear network traffic sequence is converted into a network traffic subsequence of different frequency bands, and then the spectrum energy variation of the signal can be found to be abnormal in the network. Traffic, and thus the specific location where traffic anomalies occur.

3. Data Sources

Establish a network simulation environment, collect the value of the MIB variable ipForwDatagrams of a random router in the network, assuming the observation value is $\{Z_i\}$, where i=1, 2, ..., 1830, the sampling interval is 1 second, at time At 1800, a server near the router crashed and did not work at all. Perform anomaly detection, calculate the anomaly detection statistic λ of each point in the sequence, and obtain a statistic sequence {λ}, where $\{\lambda^+\}$ and $\{\lambda^-\}$ are respectively a sequence of positive and negative values in $\{\lambda\}$, respectively, and their numbers are respectively m and n, σ^+ and σ^- respectively represent their standard deviation, is

$$\overline{\lambda^+} = \frac{\sum_{i=1}^{m}\lambda_i^+}{\mathrm{m}} \qquad \sigma^+ = \sqrt{\frac{1}{\mathrm{m}-1}\left[\sum_{i=1}^{m}(\lambda_i^+ - \overline{\lambda^+})^2\right]} \tag{6}$$

$$\overline{\lambda^-} = \frac{\sum_{i=1}^{n}\lambda_{\mathrm{i}}^-}{\mathrm{n}} \qquad \sigma^- = \sqrt{\frac{1}{\mathrm{n}-1}\left[\sum_{i=1}^{n}(\lambda_i^- - \overline{\lambda^-})^2\right]} \tag{7}$$

Now define the allowable range of the statistic. If the average is $\overline{\lambda}$ and the standard deviation is σ, then the allowable range of the statistic, that is, the interval where the statistic is normal, is: $\left[\overline{\lambda^-}-3\sigma^-, \overline{\lambda^+}+3\sigma^-\right]$, which is the abnormal point. Based on this, the allowable range of the ipForw-Daragrams statistic is: [-1, 32, 3, 11]. From this, we can get the abnormal point in $\{Z_{\mathrm{i}}\}$, and detect the moral abnormality by GLR detection method.

4. Results and Discussion

This paper is based on PPCA model and VB rank inference algorithm to detect the detection rate of these two algorithms under a variety of strong attacks. The normalized ratio of the average attack traffic transmission rate to the link maximum bandwidth during attack strength. The false positive rate is the ratio of false positive samples to the total number of alarms. As shown in Figure 1, the difference between the false positive rate and the attack strength of the two algorithms is observed. The algorithm based on the PPCA model effectively reduces the false positive rate. As the attack strength increases, the false alarm rate decreases significantly, and the average rate of decline reaches about 32%, which can effectively suppress false positives.

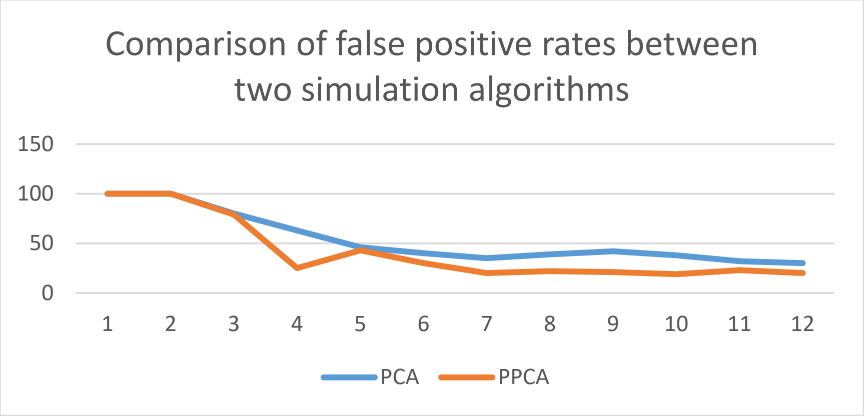

Figure 1. Comparison of false positive rates between two simulation algorithms

The detection rate is the ratio of the number of correctly detected attack samples to the total number of attacks. The curve of the detection rate of the two algorithms as a function of attack intensity is shown in Fig. 2. Obviously, with the increase of attack intensity, the detection algorithm based on PCA model and the detection algorithm based on PPCA model have obvious improvement in detection rate, and the improvement rate of the two is similar, so the performance is close. The main reasons for the weak detection rate are:

The distributed nature of the attack. Since Stuxnet is a distributed attack, the detection of distributed attacks depends on the degree of aggregation of the attack traffic of each branch by the detection algorithm. Therefore, the detection rate of the algorithm with high degree of aggregation is high. The detection rate of this paper is the same as that of the principal component analysis anomaly detection algorithm, which indicates that the algorithm can not effectively improve the detection rate while significantly reducing the false alarm rate.

The impact characteristics of industrial network services. Industrial networks often have impulsive background traffic, that is, in a very short time, the traffic is aggregated from each branch node to the control center node or the data center node. At this time, the main component of the background traffic will increase sharply and then decrease sharply, showing impact. Features. The detection algorithm cannot separate the attack traffic from the impact background during the generation of the impact background traffic, so the detection rate is not increased.

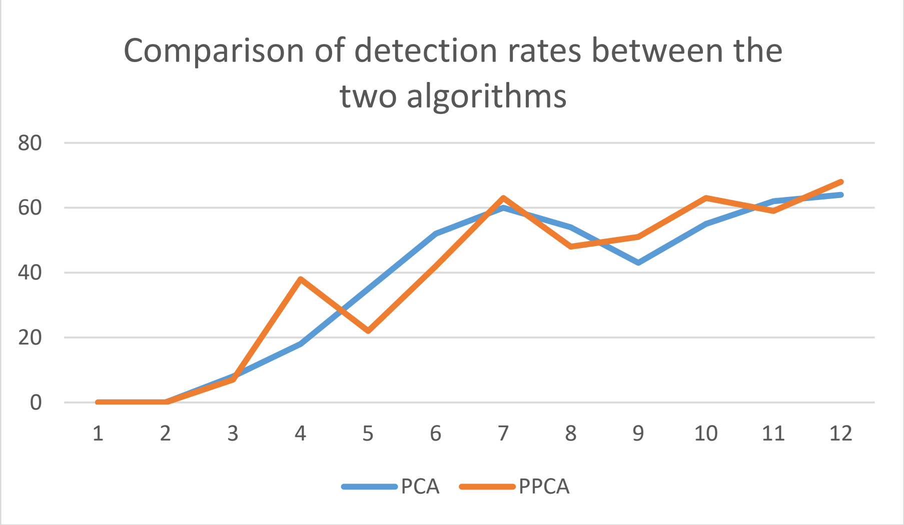

Figure 2. Comparison of detection rates between the two algorithms

The ROC curve is used to reflect the performance of the algorithm. Due to the performance limitations of the device, only test samples with sensitivity around 0%-80% can be obtained. It can be seen from the ROC curve of some samples that the algorithm based on the PPCA model has a larger area, so its performance is significantly better than PCA. Model algorithm. In this paper, the detection algorithm based on PPCA model and its rank inference is effective in reducing the false positive rate.

5. Conclusion

With the increasing use of network development, the use of the network has become more and more popular, the network scale has been continuously expanded, various new network technologies have emerged, and network equipment has become more diverse. Today, with the rapid development of the network, the rapid development of the network has also brought about an increase in the opportunities for problems in the network. It is more difficult to find the root cause of the problem, and the problems will be further spread and other issues. In order to solve these problems and maintain and maintain the normal operation of the network, what we need to do is to strengthen the management of the network. The current management of the network mainly prompts the network from false alarms. When a certain performance parameter or traffic of the network is abnormal, a crisis prompt appears, and the user is advised to close the unsafe website or page. When the network performance or traffic is abnormal, in order to detect the abnormality as soon as possible, the abnormal phenomenon is detected in time, and the abnormality needs to be continuously detected in real time. Therefore, the detection and diagnosis of the network abnormality is very necessary. The abnormal data identification based on wireless communication in this paper is for the industrial network, detecting the data of the wireless network, and discovering and repairing the abnormal data to maintain the normal operation of the network and ensure the reliability and stability of the network.

References

[1] Yi Shengwei, Zhang Yibin, Xie Feng, et al. Security analysis of industrial control network protocol based on Peach[J]. Journal of Tsinghua University(Science and Technology), 2017(01): 53-57.

[2] Zhang Kaiyi, Chen Tieming, Yan Chun. Research Progress in Safety and Abnormal Detection of Industrial Control Systems[J]. Information Security Research, 2017, 3(7): 624-632.

[3] Han Dantao, Zhao Yanling, Yan Xiaofeng. Design of Industrial Ethernet PROFINET Safety Isolator[J]. Journal of Automation, 2017(7): 46-49.

[4] LIU Wanjun, QIN Jitao, QU Haicheng. Intrusion detection method for industrial control networks based on improved single class support vector machine[J]. Journal of Computer Applications, 2018, 38(5): 1360-1365.

[5] Anonymous. Information security inspection and reinforcement of industrial control systems in service [J]. Petrochemical Automation, 2018, 54(4): 28-33.

[6] Zhu Jianjun, An Panfeng, Wan Ming. RST-SVM intrusion detection method for abnormal behavior of industrial control network[J]. Journal of Electronic Measurement and Instrument, 2018(7):8-14.

[7] Design of APT Attack Detection System in Industrial Control Network[J]. Computer Measurement & Control, 2018, 26(10): 256-260.

[8] Anonymous. A network threat awareness system architecture based on big data and machine learning[J]. Industrial Control Computer, 2018, 31(09): 120-121.

[9] Song Weiwei, Zhou Ruikang, Lai Yingxu, et al. Research on Industrial Control Anomaly Detection Method Based on Behavior Model[J]. Computer Science, 2018, 45(1): 233-239.

[10] Chen Wanzhi, Li Dongzhe. Industrial control network intrusion detection method based on whitelist filtering and neural network[J]. Journal of Computer Applications, 2018, 38(2):363-369.

[11] Han Cunwu, Chang Shurui, Qi Qi, et al. Power and Rate Optimal Tracking for Wireless

Communication Networks with Multiple Time Delays[J]. High Technology Letters, 2017, 27(4): 303-309.

[12] Lu Wei. Research on cross-layer control method for performance of wireless sensor network in microgrid data communication[J]. Journal of Electronic Measurement and Instrument, 2018(10):15-25.

[13] Xue Xue, Wang Jianping, Sun Wei. Research on cross-layer control method of wireless sensor network performance in micro-grid data communication[J]. Journal of Electronic Measurement and Instrument, 2018(10):15-25.

[14] Zheng Dongliang, Li Shichao, Zhang Liting, et al. Industrial Wireless Communication Technology Lecture 68th Application Based on WIA-PA in Extending New Ground Engineering Software Platform[J]. Instrumentation Standardization & Metrology, 2018, No.200(34): 19-22.

[15] Duan Wenjun, Xu Zhigang, Tu Fenglian, et al. Design of Rotary Weighing Filling and Capping Machine Based on Wireless Communication[J]. Packaging Engineering, 2017, 38(9): 65-68.

[16] Cui Guimei, Li Limansi, Chen Zhihui. Application of PLC Wireless Communication and Monitoring Based on S7-200[J].Wireless Communication Technology, 2017, 26(1):10-13.

[17] Zheng Dongliang, Zhang Liting, Li Shichao, et al. Industrial Wireless Communication Technology Lecture 66th Application Based on WIA-PA in Oil Production Monitoring and Optimization Analysis of Liaohe Oilfield[J]. Instrumentation Standardization & Metrology, 2017(6): 17-20 .

[18] Wei Jing, Zhang Li, Li Shucai. Application Research of Wireless Communication Technology in Coal Transportation Program Control System of Power Plant[J]. Industrial Control Computer, 2018(2).

[19] YU Li, YANG Xiaohua, TANG Xiaobo. Research on Performance of Industrial Control MIMO-STBC Communication System Based on Blind Modulation Recognition[J]. Journal of Wuyi University(Natural Science Edition), 2018(1).

[20] Underberg L, Croonenbroeck R, Wulf A, et al. A PSSS Approach for Wireless Industrial Communication Applying Iterative Symbol Detection[J]. IEEE Transactions on Industrial Informatics, 2017, PP(99):1-1.

[21] Liu H, Feng G Z, Ji L I, et al. Research on Power Line and Wireless Channel Switching Technology Based on Analytic Hierarchy Process[J]. Electric Power Information & Communication Technology, 2017.

[22] Serizawa Y, Fujiwara R, Yano T, et al. Reliable Wireless Communication Technology of Adaptive Channel Diversity (ACD) Method based on ISA100.11a Standard[J]. IEEE Transactions on Industrial Electronics, 2017, PP(99):1-1.

[23] Saxena N, Grijalva S. Dynamic Secrets and Secret Keys Based Scheme for Securing Last Mile Smart Grid Wireless Communication[J]. IEEE Transactions on Industrial Informatics, 2017, 13(3):1.

[24] Jia W, Ming Z, Chen Z. Small Data: Effective Data Based on Big Communication Research in Social Networks[J]. Wireless Personal Communications, 2018, 99(3):1391-1404.

[25] Ma Y, Wang X, Quan Z, et al. Data-Driven Measurement of Receiver Sensitivity in Wireless Communication Systems[J]. IEEE Transactions on Communications, 2019, PP(99):1-1.