An abnormal user login behavior detection method of industrial control system based on multi-dimensional probability analysis

Wenzhe Zhang1*, Chuyang Zeng1, Yang Cao1, Zuming Qin2 and Haiguang Chen3
1 Dispatch Center of China Southern Power Grid, Guangzhou, Guangdong, 510530, China
2 Chongzuo Power Supply Bureau of Guangxi Power Grid Co., Ltd, Chongzuo, Guangxi, 532200, China
3 China Southern Power Grid Digital Power Grid Research Institute Co., Ltd, Guangzhou, Guangdong, 510530, China
*Corresponding author’s e-mail: zhangwenzhe@csg.cn

Abstract. Aiming at the poor detection performance of user abnormal login behavior in the complex environment of industrial control system, a new method based on multi-dimensional probability analysis is proposed. First of all, login time, source, destination and other login parameters are collected. Secondly, the abnormal probability of login parameters is calculated. Thirdly, the login transfer probability among multiple destination hosts is calculated. Finally, through the comprehensive analysis of the above three probabilities, an abnormal login behavior detection model based on multi-dimensional probability analysis is constructed. Experiments show that this method can effectively identify the abnormal login behavior of users caused by various types of network attacks.

1. Introduction
As a combination of automatic control system and communication network, the operation of industrial control system (ICS) is highly dependent on the data collection and exchange [1,2]. User login authentication is an usual way to ensure the cyber security of ICS data exchange between hosts. Login parameters such as account, password or source IP address are verified in the authentication[3,4]. However, the authentication can be passed by with malicious login parameters. Therefore, abnormal user login behavior detections are proposed to track the cyber attacks.

In previous work, Wu Peiji et al. [5] propose an abnormal behaviour detection methods in video sequence based on deep network models and Wang Shumin et al. [6] propose an abnormal behavior identification algorithm underground mines based on continuous density hidden Markov model. However, previous detection methods suffer from the poor accuracy because of the ICS complex environment where personal login and machinery login exist. In this paper, an abnormal user login behavior detection method of ICS based on multi-dimensional probability analysis is proposed, which effectively improves the detection accuracy.
2. Detection of user abnormal login behavior of industrial control system

2.1. Detection method

The detection method mainly includes the following functional modules:

(1) Collection module. This module is designed to collect user login time, source, destination and other login parameters through the host log similar as previous work[7,8].

(2) Login time abnormal probability calculation module. This module is designed to calculate the historical login time distribution of a specific user and the corresponding abnormal probability. Generally, no matter personal login or machinery login, the login time distribution follows certain rules. For example, a personal login usually occurs during working hours and machinery login is scheduled by a fixed program setting. The login time distribution can be described and the abnormal probability of a specific login time can be calculated.

(3) Login source abnormal probability calculation module. This module is designed to calculate the historical login source (source IP) distribution of a specific user and the corresponding abnormal probability. Generally, no matter personal login or machinery login, the login source distribution follows certain rules. For example, personal logins are usually from certain fortress machines and work stations. And a machinery login is usually from certain servers and embedded devices. The login source distribution can be described and the abnormal probability of a specific login source can be calculated.

(4) Login transfer abnormal probability calculation module. This module is designed to calculate the successive login transfer probability among multiple destination hosts (destination IP). A historical login transition matrix is constructed. Generally, no matter personal login or machinery login, the login transfer follows certain rules [9,10]. For example, a personal login to the Web interface of a industrial monitoring system probably occurs after the database of the same system. And a machinery login remains constant between several destinations with a fixed probability. Based on the historical login transition matrix, a transfer abnormal probability from previous destination to current destination can be calculated.

(5) Anomaly detection module. This module is designed to integrate the above login time, login source and login transfer abnormal probability to identify users login behavior. Simply, different weights are assigned to the three aspects above and the overall abnormal probability is calculated.

2.2. Definition and implementation

2.2.1. Abnormal probability of login time

To analyze different login periods such as working day and weekend, day and night, a login time probability density curve is presented. The statistic time step of this curve is half an hour and one day is divided into 48 steps. The number of historical hitting in each step within 60 days is accumulated as \( n_i \), and the discrete probability distribution can be expressed as formula (1):

\[
 f(i) = \frac{n_i}{60}, i = 1, 2 \cdots 48
\]

Since the login time is continuously distributed in practice, curve fitting is carried out to form a continuous probability distribution curve \( f(t) \) shown in Figure 1.
Define the abnormal probability \( p_i \) of login time as the integral value of \( f(t) \) outside the half-hour step shown in formula (2).

\[
p_i = 1 - \int_{t - 0.25h}^{t + 0.25h} f(t) dt
\]

\[\text{(2)}\]

### 2.2.2. Abnormal probability of login source

Set the IP address of a login source host as vector \( \vec{x} \). So the set of multiple login sources \( \vec{x} \) is defined as \( X \). Taking the 32-bit IPv4 address as an example, every 8 bits can be considered as an element of the vector, then \( \vec{x} = [x_1, x_2, x_3, x_4] \). The normalized Minkowski distance between two login sources \( \vec{x} \) and \( \vec{x}' \) is defined as follows:

\[
dis(\vec{x}, \vec{x}') = \left( \sum_{i=1}^{4} \left[ \frac{2^{i-1} \cdot (x_i - x_i')}{x_i + x_i'} \right]^\frac{1}{4} \right)
\]

\[\text{(3)}\]

If \( \vec{x} \) represents a new login source and \( \vec{x}' \) represents an existing login source, it can be seen from formula (3) that a greater abnormal probability comes with a greater \( dis \). Further define the abnormal probability \( p_x \) of login source shown in formula (4).

\[
p_x = \min_{x \in X} \{ dis(\vec{x}, \vec{x}') \}
\]

\[\text{(4)}\]

### 2.2.3. Abnormal probability distribution of login transfer

According to the Markov chain principle, assuming that the IP address of a current login destination is vector \( \vec{y} \) which is only related to the previous login destination, the login transfer probability is described as formula (5):

\[
t_{ji} = p(y_j | y_i)
\]

\[\text{(5)}\]

In this paper, the historical data within 60 days are selected to construct the transition matrix (6).

\[
T = \begin{bmatrix}
t_{11} & t_{12} & \cdots & t_{1n} \\
t_{21} & t_{22} & \cdots & t_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
t_{n1} & t_{n2} & \cdots & t_{nn}
\end{bmatrix}
\]

\[\text{(6)}\]
Since the transfer probability is distributed discretely, the abnormal login transfer probability is presented as formula (7).

$$p_T = 1 - t_H$$  \hspace{1cm} (7)

2.2.4. Comprehensive detection based on multi-dimensional probability analysis

A abnormal login behavior detection model based on multi-dimensional probability analysis. Assuming the weights of the three abnormal probability analysis above are $\lambda_i$, $\lambda_s$, and $\lambda_T$ respectively, the overall abnormal probability $P_E$ is described as formula (8):

$$P_E = \lambda_i \cdot p_i + \lambda_s \cdot p_s + \lambda_T \cdot p_T$$  \hspace{1cm} (8)

Meanwhile, $\lambda_i + \lambda_s + \lambda_T = 1$. If $P_E > 0.5$, a user login will be determined to be an abnormal login.

3. Experimental results

The detection method is applied to a practical ICS to verify the performance. Normal and abnormal user login behaviors are simulated simultaneously. The simulation results are shown in Table 1.

| Abnormal login behavior type | Simulation setting | Effective detection | Detection accuracy % |
|------------------------------|--------------------|---------------------|----------------------|
| Personal unintentional login | Login time, sources and destinations are generated randomly. Login sources are fixed. | √ | 100 |
| Personal non-working hours login | Existing login destinations and non-working hours are selected. | √ | 97 |
| Remote personal login | Existing login destinations and working hours are selected. Login sources are fixed. Login sources are generated randomly. | √ | 98 |
| Personal login to new destinations | Working hours are selected. Login destinations are generated randomly. Login time intervals are fixed. Login sources and login time are generated randomly. | √ | 95 |
| Remote machinery login | Login sources are fixed. Login sources and login time are generated randomly. | √ | 96 |
| Machinery login to new destinations | Login time and destinations are generated randomly. | √ | 98 |

According to the 6 abnormal login behaviors widely used in cyber attack shown in table 1, this method can achieve a detection accuracy higher than 95%. Furthermore, the simulation data of
abnormal login behaviors is added to the a set of practical ICS operation data from January to June 2021 for a long-term simulation. The detection performance curve is shown in Figure 2, where TP, TN, FP, FN represents the true positive rate, true negative rate, false positive rate and false negative rate of abnormal login behavior detection.

![Figure 2. Detection performance curve in long-term simulation](image)

It is obvious that this method can achieve low FP and FN while ensuring high TP and TN.

4. Conclusion
Because the traditional identification method of abnormal login behavior of industrial control system users has the problem of low identification accuracy, this paper studies the identification method of abnormal login behavior of industrial control system users based on multi-dimensional probability analysis method. According to the function, the industrial control system is divided into acquisition module, login time anomaly probability calculation module, login location anomaly probability calculation module, historical login transfer anomaly probability calculation module and anomaly identification module, and the historical login transfer anomaly probability distribution is calculated by Markov chain principle. According to the calculation results, an abnormal login behavior recognition model based on multidimensional probability analysis is constructed. Under the condition of recognition accuracy, this method achieves low false alarm rate and missing alarm rate, and has high applicability to industrial control system. However, this method does not consider the time consumed by the identification of users' abnormal login behavior in the research process, resulting in the reduction of identification efficiency. Therefore, in the next research, the algorithm is improved to shorten the operation time, in order to quickly and accurately identify the abnormal login behavior of industrial control system users.

Reference
[1] Shen Meng, Sang Anqi, Zhu Liehuang, et al. Detection method of abnormal transaction behavior of blockchain digital currency based on motivation analysis [J]. Journal of Computer Science, 2021,44 (01): 193-208
[2] Zheng Sida, Liang Qilin, Peng Xinxia, et al. Research on detection of abnormal power consumption behavior based on fuzzy clustering [J]. Electrical Measurement and Instrumentation, 2020,57 (19): 40-44
[3] Wang Hongnian, Su Han, Long Gang, et al. Detection method of abnormal behavior in parking lot based on behavior key sentence characteristics [J]. Computer Science, 2019,46 (10): 299-306
[4] Hu xunyin, Guan yepeng Video abnormal behavior recognition method based on 3d-lcrn [J]. Journal of Harbin Institute of Technology, 2019,51 (11): 183-193
[5] Wu Peiji, Mei Xue, He Yi, et al. Detection method of abnormal behavior in video sequence based on depth network model [J]. Progress in Laser and Optoelectronics, 2019,56 (13): 134-140
[6] Wang Shumin, Chen Wei Underground abnormal behavior recognition algorithm based on continuous density hidden Markov model [J] Journal of Jilin University (Engineering Edition), 2021,51 (03): 1067-1072
[7] Liu Jia Integrated detection method of abnormal intrusion behavior in perception layer of ship communication network [J] Ship science and technology, 2020,42 (06): 133-135
[8] Chen Guoliang, Cao Xiaoxiang Pedestrian behavior recognition in indoor fire based on mobile phone multi-source sensors [J] Journal of Tongji University (NATURAL SCIENCE EDITION), 2019,47 (03): 414-420
[9] Yin Jun, Peng Yanhong, Lu Yi, et al Prediction of abnormal behavior of enterprise information system users based on deep neural network [J] Management science, 2020,33 (01): 30-45
[10] Tao Zhiyong, Guo Jing, Liu Ying CSI efficient human behavior recognition method based on multi antenna decision [J] Computer science and exploration, 2021,15 (06): 1122-1132