Research Article

Internet of Things Information Network Security Situational Awareness Based on Machine Learning Algorithms

Lei Meng

Economics and Management School of Harbin University of Science and Technology, Harbin 150080, China
Correspondence should be addressed to Lei Meng; 1705040122@xy.dlpu.edu.cn
Received 5 May 2022; Accepted 28 June 2022; Published 21 July 2022
Academic Editor: Jiguo Yu

Copyright © 2022 Lei Meng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to accurately predict the security situation of Internet of Things information network, a research method based on machine learning algorithm for security situational awareness of Internet of Things information network is proposed. The perception result is represented by the perception model, the sample data are preprocessed based on the linear discriminant analysis method, the sample data are optimized to obtain the combined features, and then the processed data are used as the training input of the RBF neural network to find out the mapping relationship with the network situation value, so as to quantify the security posture of the system.

The results show that automatic discovery and classification of seven violations is achieved. Since the platform was launched in China Mobile, more than 65,000 suspected illegal IoT cards have been discovered, effectively monitoring and controlling the operation and behavior of IoT cards. The processing efficiency of IoT card violations has been increased by more than 20 times. After the system is put into use, the discovery of suspected illegal users and behaviors can be realized through full automation, and the process of analysis, confirmation, and disposal can be shortened to 2 hours, which effectively reduce the false alarm rate and reduce operator costs. In previous monitoring of IoT cards, the false-positive rate of illegal IoT cards was about 83%, while the false-positive rate of existing algorithms dropped to 20%. Monitoring shows that business abuse monitoring detects the highest proportion of illegal IoT cards, 59%; infractions of Internet of Things cards detected through Internet abuse monitoring accounted for 20%; the proportion of Internet of Things card with machine card separation is 8%; in the information security risk monitoring (including spam text messages and harassing phone calls), the number of illegal Internet of Things cards found is small, only 3% and 2%, respectively; other infractions, including unauthorized use in locations and user complaints, accounted for 8%. It can effectively improve the ability to discover illegal IoT cards, greatly improve the accuracy of judgment, and improve the efficiency of disposal.

1. Introduction

With the continuous application of the Internet of Things in people’s life in recent years, the market size of the national Internet of Things industry is expanding. In recent years, communication technology and computer technology have developed rapidly [1]. Software vulnerabilities, also called vulnerabilities, usually refer to the defects existing in computer system software or problems arising in the process of system use. According to the definition of CVE, an authoritative vulnerability publishing organization, a vulnerability is a computational logic error found in software and some hardware components (for example, firmware) that, when exploited by an attacker, will adversely affect one of the three elements of information security (confidentiality, integrity, availability) [2]. By exploiting the vulnerability, an attacker can obtain access to the system or network, thus performing malicious acts [3]. In view of the above background, it is necessary to study the information security management system of the Internet of Things based on
machine learning algorithm. In this environment, these traditional methods are playing a huge role in the current vulnerability mining field [4]. The goal of this system software is to build a safe and accurate indoor environmental data collection system by using the Internet of Things technology, wireless transmission technology, and machine learning technology to enable data monitoring, real-time/historical display; and to provide early warning and feedback through the analysis and processing of the collected data. It is mainly composed of the following innovations: data fault-tolerant processing, the discovery of abnormal data processing, and investigation methods. The original indoor environment collection and monitoring system is improved, and through the process of analyzing and processing the collected data, the sensor lag, abnormal return value, noise and abnormal floating state of data can be evaluated. Locate the occurrence time and location of abnormal data, exclude abnormal data, make different alarm prompts for different sensor abnormal states, and enhance the accuracy and safety of the system. (2) Machine learning algorithms can be used to improve the accuracy of detecting abnormal condition. Considering the sensor anomalies such as abnormal floating and considering the state of the unbalanced data, the accuracy of such abnormal judgment is improved by using the machine learning algorithm to increase the abnormal/norm classification. With the rise of artificial intelligence, the research based on code characteristics is increasing gradually. In view of this research problem, Cui et al. proposed that deep learning has the characteristics of automation and mass, and the combination of vulnerability mining and artificial intelligence can effectively improve the defects of traditional methods that cannot be batch processing and modularization [5]. Guo et al. proposed to combine artificial intelligence technology into each step of vulnerability mining, so as to simplify the human cost required by each step and save the time cost required by each step [6]. The final results show that the intelligent detection scheme can surpass the detection accuracy of existing commercial tools [7, 8]. At present, the management of Internet of Things cards by operators is not perfect, which not only lacks the management and display system for the detailed data of the whole Internet of Things card, including the industry information of the number location, but also lacks the restrictions on users' business according to the industry usage scenarios Internet [9] of the business, not timely find IoT card business abuse, malicious attacks, and other security risk [10]. Internet of Things information security management is the use of certain security strategy, take all possible methods and means, involving the Internet of Things information security, including network, sensors, personnel, and other orderly security management, so as to ensure the integrity, reliability, confidentiality, availability, and controllability of the whole process. Compared with other information system security management, the Internet of Things information security management involves a wide range of fields, and the content is complex. Therefore, to comprehensively protect the information security of the Internet of Things, we will not only need to develop the Internet of Things information security technology but also pay attention to the information security management of the Internet of Things. Based on the current research in this paper, based on machine learning IoT card monitoring technology of concrete using the method based on machine learning algorithm to the behavior of the Internet card for identification, including the FCM algorithm and naive Bayes algorithm, and implementation in China mobile company, the results show that implements the automated detection and classification of 7 kinds of irregularities. Since the launch of the platform on China Mobile, more than 65,000 suspected illegal IoT cards have been found, effectively monitoring and controlling the operation and behavior of IoT cards. The disposal efficiency of violations of the Internet of Things card has been improved by more than 20 times [11]. After the system is put into use, it can realize the discovery of suspected illegal users and behaviors through full automation and shorten the process of analysis, confirmation, and disposal to 2h. Effective reduction in the false alarm rate reduces the cost of operators. In the previous monitoring of the Internet of Things card, the false-positive rate of illegal Internet of Things card is about 83%, while the false-positive rate of the existing algorithm is reduced to 20%. It can effectively improve the discoverability of illegal Internet of Things cards, greatly improve the judgment accuracy, and improve the disposal efficiency [12]. The Internet of Things information security monitoring system software implemented in this article provides methods for using the Internet of Things technology, Internet technology, and machine learning technology to solve the Internet of Things information security management.

2. Methods

2.1. Internet of Things Card Monitoring Technology Based on Machine Learning. In the monitoring of illegal Internet of Things card business, some violations (such as machine card separation and continuous change of position) can be judged by simple policy rules, for example, business abuse, sending illegal information, making harassing phone calls, and other behaviors; the violations cannot be found through simple filtering rules or are not accurate. Therefore, the project adopts machine learning algorithm to identify the behaviors of the Internet of Things card, including FCM algorithm and naive Bayesian algorithm [13].

2.1.1. Use FCM to Conduct Business Type Audit. FCM algorithm is a data clustering method based on the optimization of the objective function, which can perform multi-class clustering of data. The clustering result is the membership degree of each data point to the clustering center, which is represented by a numerical value [14]. The algorithm allows the same data to belong to multiple different classes. And FCM is an unsupervised fuzzy clustering method that does not need human intervention in the process of algorithm implementation. During the use of the Internet of Things card, the business behaviors and business types of the Internet of Things card are usually different from those of normal users. A normal user’s plan starts from 58
are various websites visited by normal ordinary users, abuse violations by Internet of Things card users [17]. Therefore, for normal users, most services include SMS (currently receiving more, but sending less), MMS (currently receiving more, but sending less), traffic, call (calling and called), and value-added services. For users of the Internet of Things card, the package fee is low, and users may only promise to use one or two services. For example, the Internet of Things card installed in smart camera only needs traffic and SMS service, while the Internet of Things card installed in smart meter only needs SMS service [15]. In the training process of classification, we first prepared the business data of 100,000 normal users as a positive sample and then found the business data of 10 Internet of Things cards in different industries (10,000 samples for each industry) as a negative sample, a total of 11 categories. For the Internet of Things cards to be classified, FCM algorithm can find out the probability (fuzzy value) that each sample belongs to different categories, so we choose FCM algorithm to classify business types. For a single physical network card business users, we used vector \( x = (x_0, x_1, x_2, x_3, x_4, x_5, x_6) \), where \( x_0 \) indicates user sent messages, \( x_1 \) indicates user text messages, \( x_2 \) indicates user calls calling number, \( x_3 \) indicates the number of user calls, \( x_4 \) indicates user Internet traffic, \( x_5 \) indicates user message number, and \( x_6 \) indicates opening the user number of value-added business [16]. All the above variables are normalized to make their values [0, 1].

For each of the above categories, calculate the central \( c_j \) as shown in the following formula:

\[
c_j = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad j = 1, 2, \ldots, 11.
\] (1)

Based on formula (1), for the IoT card user \( x_i \) to be classified, the probability \( \mu_{ij} \) of the card belonging to different IoT card categories is calculated, respectively, as shown in the following formula:

\[
\mu_{ij} = \frac{1}{\sum_{k=1}^{C} \left( \| x_i - c_k \| / \| x_i - c_k \| \right)^2}.
\] (2)

where \( \| x_i - c \| \) represents the Euclidean distance between vector \( x_i \) and the center of category \( c \).

After calculating \( \mu_{ij} \) through the above methods, we set the category threshold by constantly reviewing the feedback and can find out the Internet of Things cards involved in business abuse (the use of the Internet of Things card is similar to that of normal users).

2.1.2. Use Naive Bayesian Algorithm to Conduct Online Audit of Violations. When many Internet of Things cards are opened, card operators do not restrict the business types of Internet of Things card users or do not restrict the data package of Internet of Things; card users can almost use unlimited data, which leads to the possibility of Internet abuse violations by Internet of Things card users [17]. There are various websites visited by normal ordinary users, including Taobao, Pinduoduo, JD.com, Baidu, Tencent, Youku, iQiyi, and so on. Websites visited by normal Internet of Things card users include AMAP. com (Amap), Mobi-ke.com (Mobike), Aliyun (Aliyun platform), and other Internet of Things domains. Because naive Bayes is often used in text classification and has a solid mathematical foundation and stable classification efficiency, we use naive Bayes algorithm to check and find illegal Internet access: In step 1, we find the relevant Internet access data of ordinary users and Internet of Things card users as the training set and divide the Internet access data of users into two categories \( c = \{c_p, c_f\} \), where \( c_p \) represents the domain name data of normal users and \( c_f \) represents the domain name data of Internet of Things card users. The second step is to preprocess each access domain name \( D \) and decompose it. For example, in map.baidu.com, com is deleted and map and badu are retained. The third step is to calculate the probability of occurrence of a domain name word in the domain name of the training library by using the following equation:

\[
P\left( w\mid c_j \right) = \frac{\text{count}(w, c_j)}{\sum_{w \in C} \text{count}(w, c_j)}.
\] (3)

where \( \text{count}(w, c_j) \) indicates the number of times that domain name \( w \) appears in a certain \( c_j \), and \( \sum_{w \in C} \text{count}(w, c_j) \) indicates the number of times that all domain names words appear in \( c_j \).

Step 4: calculate the probability of occurrence of a domain name for the domain name \( D \) that needs new judgment, which is given by

\[
\hat{c} = \arg \max_{\hat{c} \in C} P\left( \frac{d}{c} \right),
\] (4)

where \( P(d/c) \) indicates the probability that domain name \( D \) appears in \( c_j \). For domain name \( D \), it can be represented by the characteristics of each word in its domain name, then \( d = \{w_1, w_2, \ldots, w_n\} \). In the algorithm, we assume that the occurrence of each word is independent of each other according to naive Bayes. Therefore, (5) is transformed into (6) as follows:

\[
\arg \max_{\hat{c} \in C} P\left( \frac{d}{c} \right) \times P(C),
\] (5)

where \( P(d/c) \) indicates the probability that domain name \( D \) appears in \( c_j \). For domain name \( D \), it can be represented by the characteristics of each word in its domain name, then \( d = \{w_1, w_2, \ldots, w_n\} \). In the algorithm, we assume that the occurrence of each word is independent of each other according to naive Bayes. Therefore, (5) is transformed into (6) as follows:

\[
\arg \max_{\hat{c} \in C} P\left( \frac{d}{c} \right) \times P(C) = \arg \max_{\hat{c} \in W} \prod_{w \in W} P(w/c) \times P(C),
\] (6)

where \( P(w/c) \) is calculated according to (3). Since the probability value is small, in order to avoid underflow in the calculation process, the logarithm function \( \log \) is introduced, and (6) is transformed into (7) as follows.

\[
\arg \max_{\hat{c} \in C} (\log P(C)) + \sum_{i=1}^{n} P(w_i|C).
\] (7)

Based on naive Bayes algorithm, we can classify online domain names as well and determine the threshold value.
through experiments to find out the access domain names of normal users accessing the Internet of Things card [18].

2.1.3. Use Naive Bayesian Algorithm for SMS Classification. Similar to the Internet domain name classification, the project can also use naive Bayesian algorithm to classify the suspected illegal short messages sent by Internet of Things card users: Step 1: we find ordinary short messages and divide them into two categories: \( c = \{c_0, c_1\} \) and try to ensure that the sample number of training sets in each category is similar. The second step is to carry out word segmentation for each message and preprocess to remove data such as "of", "I," and punctuation marks and retain key word information. In the third step, we also use formula (3) to calculate the probability of occurrence of a certain Chinese word, such as "real estate," "preparing for the exam," and other words appear more frequently in advertising short messages (which are illegal short messages). In step 4, formula (7) is used to calculate the probability that each SMS falls into the normal SMS category or the illegal SMS category after word segmentation, and then the threshold value is determined to find out the cases where the Internet of Things card users are suspected of discovering illegal SMS.

The Internet of Things WPDRRC Information Security Model is a dynamic information security model for the Internet of Things, improved by the WPDDRC information security model. The IoT WPDRRC model has six links and two elements, namely early warning (W), protection (P), detection (D), response (R), recovery (R), counterattack (C), forming a dynamic closed-loop WPDDRC model adding detection (D), response (R), recovery (R), counterattack (C) to the P2DR model, while combining personnel, management, and technology to effectively ensuring the safety of the information system.

2.2. Architecture and Implementation of Internet of Things Card Security Risk Monitoring System. The Internet of Things network architecture is composed of application layer, processing layer, transmission layer, and perceptual layer. According to the current Internet of Things, information security management research is in the model stage, and this paper proposes to apply the Internet of Things information security management system software to the system design and designs the Internet of Things information security management system software. The results show that the designed Internet of Things information security management system software includes equipment management, system management, network security configuration management, encryption technology management, and data mining management. Based on the Internet of Things card monitoring technology, we have deployed the Internet of Things card security risk monitoring system in China Mobile Liaoning Company, which is used to monitor the traffic related to the Internet of Things card and find out the Internet of Things card with abnormal behavior. The following is a detailed description of the system implementation.

2.2.1. Detailed Functions of the System. (1) Internet of Things Card full traffic monitoring: for all traffic and data related to mobile Internet of Things card users, including Internet log data, DNS log data, call signaling data, suspected SMS data, user signing data, and user consumption data, select all traffic related to the Internet of Things card and perform data normalization, key field extraction, and storage. In addition, the existing technical basis of text, signaling, and URL determination can be used to build professional security data. On this basis, the industry card Internet of Things card is monitored as in Table 1.

(2) Methods for monitoring illegal Internet of Things cards: the system covers a wide range of risk analysis of the Internet of Things card business and carries out highly targeted analysis of different violations. Based on the machine learning algorithm described above, training is carried out for different business scenarios, and various abnormal business behaviors are continuously tracked and identified. User consumption location analysis: behaviors of Internet of Things card users in unauthorized areas are found. Online usage audit: the Internet of Things card users and normal users are found to have similar online behaviors based on user domain names. Separation analysis of machine and card: according to the characteristics of IMEI change of Internet of Things devices, monitor the separation of machine and card existing in Internet of Things cards. Information security risk monitoring Botnet control detection: it is found that the Internet of Things card terminal is controlled. Nuisance call detection: suspected nuisance calls made by Internet of Things card users were found. Spam message analysis: Internet of Things card users sent suspected spam messages.

2.2.2. Construction of Internet of Things Card Security Monitoring System. The system can classify the Internet of Things cards that have been determined to be illegal online. In the user visualization interface, the illegal content, user-related information, and other data of the Internet of Things card can be displayed in detail or macroscopically, and the development trend of the security monitoring system of the Internet of Things card can be found. In addition, the illegal Internet of Things card can be collected and confirmed twice offline, and the confirmed illegal card can be dealt with, and the disposal results can be updated in system. A closed-loop security monitoring and processing mechanism for the Internet of Things card is formed to build a complete security monitoring system.

2.3. Relevant Technology and Theory Introduction of the Internet of Things Information Security Management System Software. The software model of the Internet of Things information security management system based on machine learning algorithm is improved by P2DR model. Combined with the management item relationship in ISO27001 standard, with information security strategy as the control and guidance, the circular dynamic security management system of protection (P), detection (D), and response (R), in which the security measures of protection, detection, and response
provide support for the security strategy. Machine learning algorithms, an oversampling technique for synthesizing a few classes of samples, are a powerful method widely believed to successfully solve problems in all kinds of applications. Similar data by artificial generation in the few existing class samples are specifically for minority set $S_{\text{min}} \in S$, $s$ is the sample set. For each $X_i$ belonging to $S_{\text{min}}$, some special $k$-nearest neighbors are determined. K-nearest neighbor is defined as $k$ elements in $S_{\text{min}}$, which satisfies the nearest Euclidean distance of distance $X_i$ in $n$-dimensional feature space $X$. Then, the oversampling process is carried out, one of the $k$ nearest neighbors is randomly selected, the random number is multiplied by the corresponding eigenvector in the interval $[0,1]$, and finally the newly generated vector is added to $X_i$:

$$x_{\text{new}} = x_i + \delta (x_i - \bar{x}_i)$$

(8)

Among them, $X_i \in S_{\text{min}}$ is the few currently selected samples; $\bar{x}_i$ is $x_i$, one of the $k$ neighbors; $X_i$ belongs to $S_{\text{min}}$ and $\delta \in [0,1]$ is a random number. Therefore, the sampling example generated according to formula (8) is a random sample point along $X_i$ to $k$ nearest neighbor $\bar{x}_i$.

### 3. Results and Analysis

#### 3.1. Requirements Analysis of the System Software

The main design goal of the system is to design an intelligent monitoring system suitable for real-time monitoring of factories, home furnishing and warehouses by using modern Internet of Things technology, sensor technology, wireless network technology, web development technology, and other related technologies. The system can carry out data acquisition, website monitoring and control, data storage and analysis, and support the expansion of various sensors to measure environmental parameters at any time. Users can view the indoor environmental state from the computer, mobile phone, tablet, and other terminal devices and control the relevant temperature control and other execution devices. In addition, the system has added a monitoring and early warning module, which has a certain data self-inspection function. In addition to the storage and display of the collected data, it will conduct a specific analysis of the abnormal changes of the collected data. The system will conduct different feedback processing for different abnormal types, analyze the data, and make the abnormal category judgment. The abnormal data detection algorithm based on SMOTE and SVM is integrated to make a more accurate judgment of the abnormal data and improve the sensitivity, timeliness, reliability, and safety of the system. Figure 1 shows a brief design diagram of the overall software architecture of this system.

#### 3.2. Internet of Things Card Security Risk Monitoring Effect

##### 3.2.1. Overall Monitoring Results

The system has been online and applied since 2019. It monitors and analyzes Internet of Things card users, including the average daily analysis of Internet of Things user signaling and Internet log exceeds 105 GB; the cumulative analysis data exceeded 15 TB; it monitors 2.9 million active IoT card users on a daily basis. More than 400 suspected illegal Internet of Things cards were found every day, and nearly 350 violations were confirmed by manual audit every day, totaling more than 65,000. Mobile IoT card abuse accounts for more than 2%. Monitoring shows that business abuse monitoring detects the highest proportion of illegal IoT cards, 59%; infractions of Internet of Things cards detected through Internet abuse monitoring accounted for 20%; the proportion of Internet of Things card with machine card separation is 8%; in the information security risk monitoring (including spam text messages and harassing phone calls), the number of illegal Internet of Things cards found is small, only 3% and 2%, respectively; other infractions, including unauthorized use in locations and user complaints, accounted for 8%. The overall analysis of Internet of Things card abuse is shown in Figure 2. The number of Internet of Things card abuse for instant communication is the largest, more than 5000; the abuse of shopping websites and video applications also exceeded 2000 [19]. The rest is for accessing other types of websites.

Similarly, in the analysis of information security risk, the analysis of spam SMS sent by the Internet of Things card is shown in Figure 3. The number of advertising messages found in this project is the largest, followed by illegal messages, fraud messages, and virus messages.
Figure 4 shows the analysis of the separation of the Internet of Things machine card. Through the corresponding situation of IMSI and IMEI, we found that there were more than 5700 Internet of Things cards with machine-card separation, among which the single user card (IMSI number) appeared in more than five terminals for 766 times.

3.2.2. Benefits of System Application. The Internet of Things card security risk monitoring system uses the method based on machine learning to analyze and realize the automatic discovery and classification of seven types of violations. Since the launch of the platform on China Mobile, more than 65,000 suspected illegal IoT cards have been found, effectively monitoring and controlling the operation and behavior of IoT cards. The disposal efficiency of violations of the Internet of Things card has been improved by more than 20 times [20]. In the previous complaint-based monitoring process of the Internet of Things card, manual analysis was required, and the processing period was about 2 working days. After the system is put into use, it can realize the discovery of suspected illegal users and behaviors through full automation and shorten the process of analysis, confirmation, and disposal to 2 h. Effective reduction in the false alarm rate reduces the cost of operators. In the previous monitoring of the Internet of Things card, the false-positive rate of illegal Internet of Things card is about 83%, while the false-positive rate of the existing algorithm is reduced to 20%.

4. Conclusions

This paper proposes the study of Internet of Things card monitoring technology based on machine learning-based algorithm to identify the behavior of Internet of Things cards, including FCM algorithm and Naive Bayes algorithm, and is implemented in China Mobile Company. The results show the automatic discovery and classification of seven types of violations. Since the launch of the platform on China Mobile, more than 65,000 suspected illegal Internet of Things cards have been found, effectively monitoring and controlling the operation and behavior of the Internet of Things cards. The efficiency of handling violations of the Internet of Things cards has been increased by more than 20 times. After the software of the system is launched, the discovery of suspected illegal users and behaviors can be fully automated, and the process of analysis, confirmation, and disposal can be reduced to within 2 h. Effective reduction in the false alarm rate reduces the operator cost. In previous IoT card monitoring, the discrimination false alarm rate of illegal IoT cards was about 83%, while the existing
algorithm was reduced to 20%. It can effectively improve the discovery ability of illegal Internet of Things cards, greatly improve the judgment accuracy, and improve the disposal efficiency.

**Data Availability**

The data used to support the findings of this study are available from the corresponding author upon request.

**Conflicts of Interest**

The authors declare that they have no conflicts of interest.

**Acknowledgments**

This study was funded by the National Social Science Fund Project: On the MicroMechanism and Implementation Path of the Impact of Green Finance on the High-Quality Development of Real Economy, Project No. 20BJL028.

**References**


