Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9894708, 1 page
https://doi.org/10.1155/2023/9894708

Retraction

Retracted: Intrusion Detection Method Based on Deep Learning

Wireless Communications and Mobile Computing

Received 8 August 2023; Accepted 8 August 2023; Published 9 August 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] C. Tian, F. Zhang, Z. Li et al., "Intrusion Detection Method Based on Deep Learning," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 1338392, 8 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 1338392, 8 pages
https://doi.org/10.1155/2022/1338392

Research Article

Intrusion Detection Method Based on Deep Learning

Chongrui Tian,[1,2] Fengbin Zhang,[1] Zhaoxiang Li,[2] Ruidong Wang,[1] Xunhua Huang,[1] Liang Xi,[1] and Yi Zhang[2]

[1]*College of Computer Science and Technology, Harbin University of Science and Technology, China*
[2]*College of Information Engineering, East University of Heilongjiang, China*

Correspondence should be addressed to Fengbin Zhang; 1910400004@stu.hrbust.edu.cn

Received 21 March 2022; Revised 3 May 2022; Accepted 19 May 2022; Published 15 July 2022

Academic Editor: Kuruva Lakshmanna

In today's information age, the scale of the Internet is growing, the information capacity is also expanding explosively, and network security is becoming more and more important. Intrusion detection is regarded as a traditional security protection technology and is a key means to ensure the security of the network environment. Among them, the deep belief network performs well, and it can automatically learn abstract features for classification. In order to further improve the detection rate and reduce the false positive rate, it is necessary to improve the detection rate of small sample data. This paper builds an intelligent deep learning model and analysis model for intrusion detection data based on TensorFlow. By learning to identify network intrusion characteristic data, the characteristic data and model are stored in the big data storage system built by Hadoop. This algorithm has achieved good experiment result. Build a model knowledge base and an intrusion feature behavior library, use the decision tree model to automatically match the security control strategy, realize a highly intelligent security control model with self-learning ability, and solve the rapid identification of unknown intrusion behaviors. Experiments show that the algorithm can effectively improve the detection rate.

1. Introduction

The rapid development of the Internet, on the one hand, promotes the rapid development of human beings. The traditional computer security thought can no longer satisfy the growing, multidimensional, and interconnected network environment. With the continuous development of technologies such as the Internet of Things and cloud computing, as well as the arrival of the era of big data, Internet security issues around the world have become increasingly prominent. Therefore, using machine learning technology to analyze a large amount of network traffic to determine the intrusion behavior is an effective way to enhance the security of the network.

The increasingly complex network environment makes it difficult for simple machine learning methods to solve practical problems. Since the deep learning network was photographed by Professor Geoffrey Hinton of the University of Toronto in 2006, the development of deep learning technology has had a wide-ranging impact on the research on signal and information processing. Deep learning has greatly expanded the field of machine learning research and promoted the rapid development of artificial intelligence. Due to its powerful feature expression capabilities, machine learning models based on deep neural networks have made breakthroughs in speech recognition, image recognition, and natural language processing and have received more and more attention from scholars at home and abroad. Some foreign researchers have applied it to human intrusion detection: literature [1] uses a hybrid clustering and neural network method to achieve human intrusion detection; literature [2] uses clustering-based computerized data to detect zombies. On the Internet, literature [3] uses deep neural network to secure the Internet of Vehicles; literature [4] realizes human intrusion detection based on deep belief network. There are few researches on human intrusion detection based on deep neural network in China. Literature [5] expounds the application of deep neural network in big data analysis. Literature [6] uses a two-layer restricted Boltzmann machine for structural dimension reduction, using the BP neural network obtaining the optimal representation of the original data and then using SVM to identify human intrusions on the data. Reference [7]

proposes a feature selection based on information gain for the problem that the high-dimensional features of the data in the detection of abnormal human intrusions will affect the detection rate. The detection model improves the detection rate of the random forest classifier by 0.2%.

This paper builds an intelligent deep learning model and analysis model for intrusion detection data based on TensorFlow. By learning to identify network intrusion characteristic data, the characteristic data and models are stored in the big data storage system built by Hadoop. The decision tree model is used to automatically match the security control strategy, to realize a highly intelligent security control model with self-learning ability, and to solve the rapid identification of unknown intrusion behaviors.

2. Big Data Storage for Intrusion Detection Based on Hadoop

At present, the field of network and information security is facing brand-new challenges. On the one hand, with the advent of the era of big data and cloud computing, the security problem is becoming a big data problem. The network and information systems of enterprises and organizations produce a large amount of security data every day and produce it faster and faster. On the other hand, the state, enterprises, and organizations face a severe security situation in cyberspace, and the attacks and threats to be dealt with are becoming increasingly complex. These threats are characterized by strong concealment, long incubation period, and strong sustainability. In the face of these new challenges [8, 9], the limitations of the existing security management platforms without effective solution.

In order to fully analyze the way map works, the input data of the following examples are mainly considered in designing and building Hadoop-based intrusion detection big data analysis (page length, some unused columns have been removed and indicated by ellipses): 006701199099999 2020102619230212234051507004...9999999N9+00001+9999 9999999...00430119909999920201026192302122340515120 04...9999999N9+00221+99999999999...004301199099999202 0102619230212234051518004...9999999N9-00111+999999 99999...00430126509999920201025192302122340324120 04...0500001N9+01111+99999999999...0043012650999992 02010251923021223403241800 4...0500001N9+00781+99 999999999...

These rows represent the access map function structure of the data stored in the Hadoop by key/value pairs, as follows:

(0, 006701199099999202010261923021223405150700 4... 9999999N9+00001+99999999999...)

(106, 004301199099999202010261923021223405151200 4...9999999N9+00221+99999999999...)

(212, 004301199099999202010261923021223405151800 4...9999999N9-00111+99999999999...)

(318, 004301265099999202010251923021223403241200 4...0500001N9+01111+99999999999...)

(424, 004301265099999202010251923021223403241800 4...0500001N9+00781+99999999999...)

The corresponding key in Hadoop is the line offset in the file, which is very overlooked in the map function in the process of designing and building a Hadoop-based intrusion detection big data analysis. The function of map function only extracts network exception data, security decision data, and security experts' experience stored in Hadoop database and sends it as output (the timestamp has been interpreted as an integer). This relationship is as follows:

(2020102619230212234, 0)
(2020102619230212234, 22)
(2020102619230212234, ?11)
(2020102519230212234, 111)
(2020102519230212234, 78)

The output of the map function is first processed by the MapReduce framework and then sent to the reduced function. This process sorts and groups key/value pairs by key to [10]. Therefore, continuing with examples in the design and construction of a Hadoop-based intrusion detection big data analysis, the reduced function sees the following input:

(2020102519230212234, [111, 78])
(2020102619230212234, [0, 22, 11])

All are timestamped with a series of feature data store ID. All reduce functions must now repeat this list and identify the relevant storage ID [10, 11] required for the intrusion detection analysis algorithm:

(2020102519230212234, 111)
(2020102619230212234, 22)

This is the final output: the intrusion detection algorithm in the feature data record in the Hadoop storage system matches and links the highest feature data based on the intrusion detection big data built by Hadoop. The whole data flow is shown in Figure 1. At the bottom of the graph is the system-level Unix pipeline, simulating the entire MapReduce process, the content in the design, and construction of Hadoop-based intrusion detection big data analysis in the detailed design and implementation of Hadoop intrusion detection algorithm.

After implementing the design principle of big data store access map, we can implement java-based code [12, 13] in the system. In the process of designing and building a Hadoop-based intrusion detection big data analysis, three functional interfaces are required: a map function, a reduce function, and some code to run the job. The map function is implemented by a Mapper interface where a map () method is declared, and it is reconstructed. The following core code implements the implementation [14, 15] of the map function in the process of designing and constructing the big data analysis of intrusion detection based on Hadoop: (1) the Mapper interface for the highest feature data sample inspired by Hadoop and related notes, using Java as a tool [16, 17].

The Mapper interface designed above for intrusion detection feature data access is a generic type that has 4 formal parameter types, which specify the input key, input key, output value, and type key of output value of the map function. For the current example, the input key is a long integer offset, the input value is a line of text, the output key is a timestamp, and the output value is a characteristic data (integer). Hadoop specifies its own set of basic types that can be used for network sequence optimization instead of using the built-in Java type. These can all be done in the org.apache.hadoop. Found in the io package. Type LongWritable (a Java), a type Text (a Java String), and a IntWritable (Integer) [18] are now used in the design and construction of Hadoop-based big data analysis

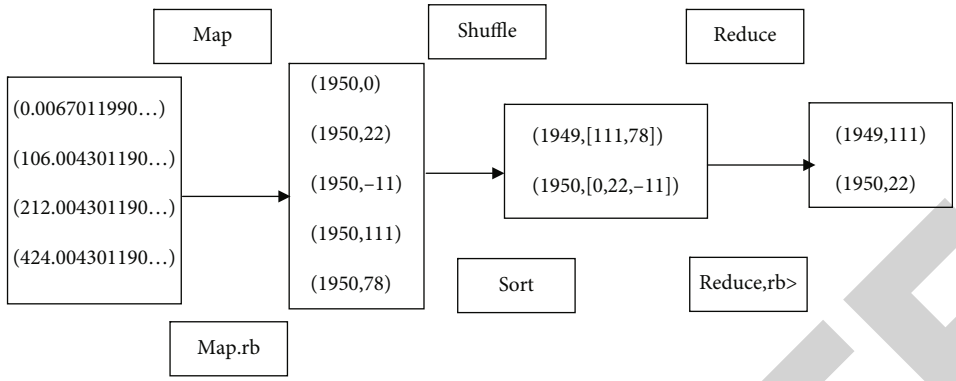

Figure 1: The MapReduce data access flow.

of Java. The map () method requires to pass in a key and a value. In the process of designing and building a Hadoop-based intrusion detection big data analysis, convert a Text value containing the input line of a Java string into a String type and then use its substring() method to extract the columns of interest in the process of designing and building a Hadoop-based intrusion detection big data analysis. The map () method also provides an OutputCollector instance to write to the output. In this case, write a timestamp as a Text object during the design and construction of Hadoop-based intrusion detection big data analysis (because only one key is used in the design and construction of Hadoop-based intrusion detection big data analysis) and wrap the timestamp [19] with IntWritable type. In the process of designing and building a Hadoop-based intrusion detection big data analysis, the output record [20] is written only after the feature data is displayed, and its quality code represents the correct feature data store ID.

The reduced function is also defined when using Reducer, as defined in the deep learning-based intrusion detection big data store designed and built in this paper as follows: definition of Reducer intrusion detection feature big data access interface of the highest feature data sample [16, 17].

Similarly, the four formal parameter types are used to specify the input and output types of the reduced function. The input type of the reduced function must match the output type of the map function: the Text type and the IntWritable type. In this case, the output type of reduce function is Text and IntWritable, the former type of time stamp, and the latter type of the highest feature data, where all feature data are traversed during the design and construction of Hadoop-based intrusion detection big data analysis, and each record is compared until the best feature data match is found [21].

3. Design of Big Data Security Control Model

3.1. Algorithmic Optimization Method. The algorithm optimization method is mainly divided into two processes: data preparation stage and intrusion detection stage. In the data monitoring stage, the system mainly collects the historical behavior data of the network users and simultaneously preprocesses the extracted user behavior data. In the process of preprocessing, the data is mainly analyzed for data cleaning and data sorting. The features of the processed data were also extracted. Store the extracted feature information into the regulatory database. Then, when used for the comparative analysis of the subsequent tests. In the process of comparison, if the risk can be found, implement the risk warning processing. Otherwise, continue with the testing. Its entire intrusion detection process is based on network big data security control [22].

3.1.1. Abnormal Traffic in the Network Detection. In the design process of the traffic anomaly detection model, based on the abnormal points of the traditional model, the snort rules are encoded and applied to the snort in the network intrusion detection system, so that snort has the function of traffic anomaly detection. This is the basic idea of the traffic anomaly detection model in the intrusion detection system based on deep neural network [23]. Based on the above ideas, the flow anomaly detection model is shown in the figure [24].

3.1.2. Design of Anomaly Model for Network Protocol Abuse Detection. Anomaly detection for network protocol abuse is primarily based on network anomaly detection, rather than just examining a single network request or response. During the detection process, complex network intrusion behaviors such as multiple attacks can be detected according to the protocol status information of the network data stream [25].

Protocol abuse anomaly detection first needs to analyze the protocol. After analyzing and identifying the protocol type, we need to add some protocol status information and then use the data flow of the entire session as the detection object. Providing a complete FTP session on the network consists of (1) establishing a TCP connection to the server on a TCP port; (2) authentication, sending the username and password through the FTP port, or some FTP that allows anonymous login; (3) the client, if the client requests a temporary port and a server link for related data transfer; (4) port TCP link is required after the FTP session ends. The entire session process based on the FTP protocol is composed of a series of ordered protocol packets. The protocol anomaly detection model requires the entire session data flow as the inspection object. According to the RFC draft, all network connectivity protocols have certain status information. Certain events must occur at certain times. Therefore, based on the above analysis, a protocol anomaly

detection model can be built in the state machine; each state is associated with a state to a process, and the state is pointed to a list of attributes and features of the system once. Protocol anomalies are designed and constructed based on big data analysis. During the detection process, the state can be judged and detected. Based on the above analysis, a network protocol model for anomaly detection is proposed [16].

Based on the above traffic anomaly and network protocol anomaly detection models, an efficient deep learning algorithm suitable for processing large amounts of data in network intrusion detection systems is proposed, based on this algorithm, traditional traffic anomaly detection and network protocol anomaly detection, and a new intrusion world model.

3.2. Detailed Design of the Optimization Model. In the above model, the core steps include data preprocessing, based on big data deep learning processing algorithm association mining, build decision tree and through the detection model detection four core process and through the detection model detection mainly using the previous traffic anomaly detection model and network protocol abnormal detection model two models for detection. The design of the other three core processes is detailed below for the following [17].

3.2.1. Data Preprocessing. In the process of network invasion behavior detection, the event formats of various network security behaviors are different. Therefore, in the process of using deep learning algorithm, it is necessary to standardize and format the various types of time collected from the network time database, and store them in the form of unified standard data. The process table of the data preprocessing is given in Table 1.

At the same time, in the research of this paper, peer experts in the field of network security big data analysis will be used as the knowledge carrier, using the professional knowledge mastered by peer experts, and the knowledge representation method in knowledge set theory will be adopted to describe and represent the knowledge of experts in [17]. The specific mathematical model is shown with

$$KS_E = (N_E, C, V) = \begin{bmatrix} N_E & c_1 & v_1 \\ & c_2 & v_2 \\ & c_n & v_n \end{bmatrix}. \quad (1)$$

N_E: in the above model, the peer expert is represented, and C is the feature set of the peer experts, which can be represented through a one-dimensional vector, as follows:

$$C = \{c_1, c_2, \cdots, c_n\}. \quad (2)$$

N_E: V represents the set of quantity values about C, which can be represented by a one-dimensional vector model. The specific mathematical model is with

$$V = \{v_1, v_2, \cdots, v_n\}. \quad (3)$$

Based on the above model, if applied to specific examples, the specific description of the information of peer experts in the field of computer network security is as follows:

Based on the above methods, in the specific processing, in the process of designing and constructing intrusion detection big data analysis, the entire expert knowledge system can be described and modeled through multidimensional vector patterns [26].

In the analysis in the field of network security big data analysis, the network security information strategy is mainly used as the knowledge carrier, and the knowledge contained in the network security big data is called security strategy knowledge. According to the knowledge set method, the security strategy knowledge can be expressed by

$$KS_p = (N_p, C, V) = \begin{bmatrix} N_p & c_1 & v_1 \\ & c_2 & v_2 \\ & c_n & v_n \end{bmatrix}. \quad (4)$$

In the above model, where the network security big data is represented, C represents the feature collection of the whole network security big data, which is a one-dimensional vector, represented as with Formula (2).

N_p: V represents the set of values about C, which is also a one-dimensional vector model. Its specific representation can be represented by Formula (3).

According to the above security policy knowledge model, taking the actual network security big data analysis field as an example, the knowledge model can be expressed as the following model.

In the knowledge representation of security policy, it is similar to the expert's knowledge representation method and structure. Using the above representation, when analyzing the matching degree between experts and projects, matching analysis can be carried out from the knowledge level to obtain the relevant similarity [27].

In the construction of the above-mentioned expert and security policy knowledge model, the knowledge of experts and security policies can be completed. In the process of expert selection, the essence is to judge the similarity between experts and security policy knowledge. Based on the similarity clustering processing, a batch of matches is finally obtained. Degree experts for project-related reviews, the specific implementation methods, and principles are as follows:

(1) Similarity calculation

This paper assumes that the knowledge set in the field of network security big data analysis is shown in

$$KS_1 = \{KS_{11}, KS_{12}, \cdots KS_{1m}\}. \quad (5)$$

The collection of expert knowledge of peer experts is shown with

$$KS_2 = \{KS_{21}, KS_{22}, \cdots KS_{2m}\}. \quad (6)$$

The similarity can be calculated according to the similarity definition of knowledge in the knowledge set theory. The specific calculation model is as follows:

Table 1: Data preprocessing procedure table.

Behavioral path	Behavioral path	Behavioral path
Source port and destination port of the change behavior	Source port and destination port of the change behavior	Source port and destination port of the change behavior
Alphabetic string	Alphabetic string	Alphabetic string
Behavioral object	Behavioral object	Behavioral object
Hostname of the behavior	Hostname of the behavior	Hostname of the behavior
Alphabetic string	Alphabetic string	Alphabetic string

$$KS_1 \square KS_2 = \begin{bmatrix} KS_{11} \\ KS_{12} \\ \vdots \\ KS_{1m} \end{bmatrix} \square [KS_{21} KS_{22} \cdots KS_{2n}]. \quad (7)$$

In the above model,

$$KS_{1i} \square KS_{2j} = KS_{1i} \square \begin{bmatrix} f_1(c_1) \\ f_2(c_2) \\ \vdots \\ f_n(c_n) \end{bmatrix} (i-1, 2, \cdots, m\,; j = 1, 2, \cdots, n). \quad (8)$$

In the model, { } is a set of related features of security policy knowledge and expert knowledge, representing the corresponding correlation function [28]

$$\{f_1(x), f_2(x), \cdots, f_k(x)\}. \quad (9)$$

Based on the above analysis, the similarity between projects and experts can be calculated through the following mathematical model:

$$KS_R\left(KS_{1i} \square KS_{2j}\right) = \sum_{i=1}^{k} f_i(w_i). \quad (10)$$

In the above model, the weight of the corresponding eigenvector $\{c_i\}$ is indicated.

On the basis of the above similarity calculation, the K-mean algorithm can be further used to cluster the experts in the expert set according to the security policy knowledge system. Suppose the set M represents the knowledge set of all experts in the expert database, where $(n = 1, 2, 3 \cdots, N)$, N represents the number of experts in the expert database. The similarity of each expert evaluated item can be represented as the following model [29].

$$\{M_1, M_2, \cdots, \mathrm{M}_n\} = KS_1\left(KS_{1i} \square KS_{2j}\right)_n. \quad (11)$$

Based on the above analysis, the similarity is used as the similarity of cluster analysis, and K classes in a given expert database are found by the K-mean method. The center of the class is the mean class based on all the values in the expert database, describing each the small size of the class. The calculation model is as follows.

$$J = \sum_{k=1}^{k} \sum_{i=1}^{n} \square x_i - u_k \square^2. \quad (12)$$

Combining the least squares and the Lagrangian principle, the cluster center is the average of the data points in the corresponding category, and in order to make the algorithm converge, the final cluster center should be kept as constant as possible during the iteration process. The closest experts to the resulting cluster center represent the most matching experts. In the processing process, the system can also select N experts near the cluster center to form an expert group for evaluation. The clustering analysis process of the similarity of K-mean algorithm is as follows:

(1) According to the similarity method, complete the similarity calculation of all experts relative to the security policy knowledge model to be evaluated, and then, in the data space composed of similarity, K samples in the data space are randomly selected for initialization, and each object represents a cluster center for processing

(2) For all the similarity parameters of the description experts and items in the sample, the Euclidean distance is further divided into the class corresponding to the nearest cluster center according to the nearest distance criterion

(3) Update the cluster center, and take the mean value of all the objects in each category as the cluster center of the category, to calculate the value of the target function

(4) Determine whether the value of the cluster center and the target function has changed, if the values of the cluster centre and the target function have not changed, the output result is the same; if they have changed, return to algorithm (2) and resume the iterative analysis

(5) Generation of association rules based on big data deep learning processing algorithm

Based on the basis of the above field experts and security strategy analysis model, using the deep learning algorithm on the data table based on big data analysis technology, the bull

association model was used in the design process of this paper. Process each user behavior in the network as a transaction during network intrusion detection [30], and collect a large number of user behavior in the network to build a transaction database. The behavior of each user in the database consists of five fields, namely, time of the act, behavioral agent, behavioral object, and behavioral path. The user's behavior is marked with a unique flag ID. By executing the network big data security control, form the following association rules (Table 2).

Based on the above association rules and the previous experts and security strategy knowledge model, on this basis, it can be combined with decision tree model, using experts and security strategy knowledge model, scanning the data training set, using the correlation rules for vertical compression [31], and getting the preprocessing training set as shown in Table 3.

Finally, we are compressed by the clustering rule for which rules of count is less than ∗ 018, and finally, we get the compressed training sample as shown in Table 4.

3.3. Experimental Results and Analysis. To verify the effectiveness of the big data security control model for intrusion detection, simulation experiments are performed. The experimental environment is: the CPU is C eleron (R)2.53 GHz, the memory is 512 M, the database management system is Oracle 10 g, the development language is Java, and the development environment is MyEclipse.

The experiment selected 10000 data of network intrusion detection and 5 different types of credit services. The service data of this experiment was exported from the database system through the program, and the data was preprocessed before entering the intrusion detection big data security control model experiment, and the network abnormal data was normalized. The experiment is divided into two groups to test the improved intrusion detection big data security control model to verify the feasibility of the network intrusion detection mining model based on the intrusion detection big data security control model. When the minimum support sup_min is fixed at 1.5%, compare the execution efficiency of the traditional RBAC security control model algorithm (Algorithm 1), FA security control model algorithm (Algorithm 2), and the improved intrusion detection big data security control model algorithm (Algorithm 3) proposed in this paper for different number of transactions.

Three algorithm execution times will increase with the number of transactions, but the algorithm three due to the preprocessing reduce about 47% candidate set, then the algorithm one and algorithm two growth is much slower, so the number of mining transactions (tens of thousands of or even hundreds of thousands, millions) improved intrusion detection big data security control model superiority will be more obvious.

When the number of transactions is fixed at 8000, compare the execution efficiency of the RBAC security control model algorithm (Algorithm 1) and the FA security control model algorithm (Algorithm 2) and Algorithm 3 for different minimum support sup_min. Minimum support sup_min, the longer the three algorithms execute. In the experiment, when sup_min decreases from 0.5% to 0.3%, the execution time of the algorithm based on RBAC security control model increases greatly, the growth rate of FA security control model algorithm is slower than the original algorithm, while the execution time of the big data security control model of intrusion detection algorithm is the slowest. In this set of experiments, the intrusion detection big data security control model algorithm designed in this flat paper reduces the generation of candidate sets by about 52%.

Table 2: Association rule results.

Path NYX Z E N G, L I U-> principal CDN	Support	Support
Principal Bery-> path DAM	24.5	98.3
Principal LER-> path WWX	25.3	91.1
Path NYX delete sale-> principal KXL	26.3	92.3
Principal Akry-> path EKR	21.2	95.2
Path KBD read sale-> principal DOM	22.4	93.4
Path MXC delete sale-> principal DOM	23.7	91.4
Principal DOM path MXC-> read sale	25.9	93.2
Rule	24.1	92.4

4. Limitations and Disadvantages

Although intrusion detection based on deep learning has advantages in encryption attack detection and zero-day attack detection compared with traditional misuse detection, deep learning technology is still unable to achieve a wide range of applications in commercial intrusion detection systems. The important reason is that the current research on human intrusion detection based on deep learning is only carried out on good data sets, and some problems existing in the real environment cannot be effectively solved.

4.1. Identification of Short Flow. A short flow refers to a network data flow with fewer data packets. When a flow contains fewer data packets, it is difficult to obtain effective flow characteristics based on these data packets. For statistical characteristics, it is necessary to obtain sufficient flow characteristics in a sufficient number of data packets. The statistical data is only meaningful in this case. The statistical characteristics of short flow often contain a large number of null values or have strong randomness, and it is difficult to express the behavior pattern of the flow.

4.2. Detection of Stronger Encryption Protocols. For traffic using stronger encryption protocols, such as traffic using encryption protocols such as QUICTLS1.3, there is not yet a good detection method.

4.3. Traffic Behavior Characteristics Vary over Time and Space. The behavioral characteristics of network traffic will change with time and space. For example, the behavioral characteristics of normal traffic collected in schools will be different from the normal traffic behavioral characteristics of companies.

4.4. Unbalanced Data Volume. In the process of deep learning, a good data set plays a crucial role in the training of the model. In the process of training the training set, enough training

Table 3: Preconditioning training set.

ID	Behavior name (A)	Time of the act (B)	Behavioral agent (C)	Behavioral object (D)	Behavioral path (E)	Risk judgement
1	Select	Update	Update	Update	0	1
2	0	Select	0	0	Update	1
3	0	Update	Update	0	0	0
4	Insert	Insert	Insert	Insert	Select	0
...	...	...	...	...	...	...

Table 4: Compressed training sample data.

ID	Behavior name (A)	Time of the act (B)	Behavioral agent (C)	Behavioral object (D)	Behavioral path (E)	Risk judgement
1	Select	Update	Update	Update	0	1
4	Insert	Insert	Insert	Insert	Select	0
...	...	...	...	...	...	...

samples are often needed to complete the training of various parameters in the neural network.

5. Conclusion

This paper completes the network traffic big data storage and access performance analysis based on Hadoop storage, firstly, by analyzing the big data security control model requirements for intrusion detection, on the basis of the traditional invasion detection, introducing a decision tree analysis model, through a deep learning model built based on the TensorFlow platform to perform the feature analysis and network feature detection and to identify the network characteristic behavior, using the decision tree model to associate the security control policy and the security expert decision, and finally, building a Hadoop stored network traffic big data storage control model and effective control of its access performance.

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 no conflict of interest is associated with this study.

Acknowledgments

The authors are thankful to the higher authorities for the facilities provided. The research was funded within the project entitled: "New Network Technology and Information Security Research Team (HDFKYTD202101)", being a part of Strategic Research Program "Research on multi-modal active sensing and smart operation technology of agent and verification of typical scenes" supported by guiding projects of Key R & D Plans in Heilongjiang Province.

References

[1] T. Ma, Y. Yu, F. Wang, Q. Zhang, and X. Chen, "A hybrid methodologies for intrusion detection based deep neural network with support vector machine and clustering technique," *Frontier Computing. FC 2016. Lecture Notes in Electrical Engineering*, N. Yen and J. Hung, Eds., vol. 422, 2018.

[2] O. Y. Al-Jarrah, O. Alhussein, P. D. Yoo, S. Muhaidat, K. Taha, and K. Kim, "Data randomization and cluster-based partitioning for botnet intrusion detection," *IEEE transactions on cybernetics*, vol. 46, no. 8, pp. 1796–1806, 2016.

[3] K. Min-Joo and K. Je-Won, "Intrusion detection system using deep neural network for in-vehicle network security," *Plos One*, vol. 11, no. 6, article e0155781, 2016.

[4] K. Raza and S. H. Adil, "Intrusion detection using deep belief network," *Mehran University Research Journal of Engineering and Technology*, vol. 33, no. 4, pp. 485–491, 2014.

[5] Z. Lei and Z. Hu, "Infinite deep neural network method for big data analysis," *Computer Research and Development*, 2016.

[6] F. Qu, J. Zhang, Z. Shao, and S. Qi, "Intrusion detection model based on deep belief network," in *Proceedings of the 2017 VI International Conference on Network, Communication and Computing*, 2017.

[7] L. Rujun, J. Bin, and X. Yang, *Computer Application of Anomaly Intrusion Detection Model Based on Information Gain Feature Selection*, 2016.

[8] D. Y. Kim, Y. J. Woo, K. Kang, and G. H. Yoon, "Failure diagnosis system using a new nonlinear mapping augmentation approach for deep learning algorithm," *Mechanical Systems and Signal Processing*, vol. 172, p. 108914, 2022.

[9] G. Niu, X. Li, X. Wan et al., "Dynamic optimization of wastewater treatment process based on novel multi-objective ant lion optimization and deep learning algorithm," *Journal of Cleaner Production*, vol. 345, p. 131140, 2022.

[10] Y. Essam, Y. F. Huang, A. H. Birima, A. N. Ahmed, and A. El-Shafie, "Predicting suspended sediment load in Peninsular Malaysia using support vector machine and deep learning algorithms," *Scientific Reports*, vol. 12, no. 1, pp. 302–302, 2022.

[11] K. Zhou, Z. Zhang, R. Yuan, and E. Chen, "A deep learning algorithm for fast motion video sequences based on improved codebook model," *Neural Computing and Applications*, pp. 1–16, 2022.

[12] H. Yan, M. Lihua, L. Shen et al., "A short-term wind speed prediction method utilizing novel hybrid deep learning algorithms to correct numerical weather forecasting," *Applied Energy*, vol. 312, article 118777, 2022.

[13] Y. Wang, Y. Han, Y. Wu, E. Korkina, Z. Zhou, and V. Gagarin, "An occupant-centric adaptive façade based on real-time and contactless glare and thermal discomfort estimation using deep learning algorithm," *Building and Environment*, vol. 214, p. 108907, 2022.

[14] C. P. Lau, W. Ma, K. Y. Law et al., "Development of deep learning algorithms to discriminate giant cell tumors of bone from adjacent normal tissues by confocal Raman spectroscopy," *The Analyst*, vol. 147, no. 7, pp. 1425–1439, 2022.

[15] M. Giuseppe, C. Mattia, B. Andrea et al., "Diagnostic performance of deep learning algorithm for analysis of computed tomography myocardial perfusion," *European journal of nuclear medicine and molecular imaging*, 2022.

[16] K. Tamai, H. Terai, M. Hoshino et al., "A deep learning algorithm to identify cervical ossification of posterior longitudinal ligaments on radiography," *Scientific Reports*, vol. 12, no. 1, pp. 2113–2113, 2022.

[17] S. Lin, "Study on the correlation between linguistic complexity and audience recognition in college English speech contests," *4th International Workshop on Education Reform and Social Sciences (ERSS 2021)*, pp. 287–294, 2021.

[18] H. Abdel-Jaber, D. Devassy, A. Al Salam, L. Hidaytallah, and M. EL-Amir, "A review of deep learning algorithms and their applications in healthcare," *Algorithms*, vol. 15, no. 2, p. 71, 2022.

[19] S. D. Bala, M. K. K. Rony, K. Islam et al., "Weather and Covid-19 outbreak correlation in Dhaka District, Bangladesh," *Science Progress and Research (SPR)*, vol. 1, no. 4, pp. 171–175, 2021.

[20] G. Orlando, D. Raimondi, R. Duran-Romaña, Y. Moreau, J. Schymkowitz, and F. Rousseau, "PyUUL provides an interface between biological structures and deep learning algorithms," *Nature Communications*, vol. 13, no. 1, pp. 961–961, 2022.

[21] P. Kumar, J. Kuttippurath, and A. Mitra, "Causal discovery of drivers of surface ozone variability in Antarctica using a deep learning algorithm," *Environmental science: Processes & impacts*, vol. 24, no. 3, pp. 447–459, 2022.

[22] Q. Hong, S. Bo, Y. Guo, Z. Yang, L. Jun, and F. Wei, "A parallel deep learning algorithm with applications in process monitoring and fault prediction," *Computers and Electrical Engineering*, vol. 99, article 107724, 2022.

[23] I. Syem, K. Naimul, and K. Sri, "Comprehending the impact of deep learning algorithms on optimizing for recurring impediments associated with stress prediction using ECG data through statistical analysis," *Biomedical Signal Processing and Control*, vol. 74, p. 103484, 2022.

[24] S. Lee, K. Seokhyeon, and M. Sungkon, "Development of a car-free street mapping model using an integrated system with unmanned aerial vehicles, aerial mapping cameras, and a deep learning algorithm," *Journal of Computing in Civil Engineering*, vol. 36, no. 3, 2022.

[25] R. Singh and A. Bahadur, "Gender differences in spirituality and subjective well-being among working couples in Indian society," *Science Progress and Research (SPR)*, vol. 1, no. 3, pp. 120–126, 2021.

[26] Y. Bin and D. Mandal, "English teaching practice based on artificial intelligence technology," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 3, pp. 3381–3391, 2019.

[27] Y. Wang, "On college English teaching in the new era——a cross-cultural perspective," *2021 Northeast Asia International Symposium on Linguistics, Literature and Teaching*, pp. 229–234, 2021.

[28] H. Yu and L. Shubo, "Implementation and innovation of college English smart education under big data environment," *The 2021 Northeast Asia International Symposium on Linguistics, Literature and Teaching*, pp. 235–240, 2021.

[29] G. Thippa Reddy, M. P. K. Reddy, K. Lakshmanna, D. S. Rajput, R. Kaluri, and G. Srivastava, "Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis," in *Evolutionary Intelligence*, vol. 13, no. 2pp. 185–196, Springer, 2020.

[30] K. Vishnu Vandana and D. S. Rajput, "A review on the significance of machine leaning for data analysis in big data," *Jordanian Journal of Computers and Information Technology*, vol. 6, no. 1, pp. 41–57, 2020.

[31] R. M. Swarna Priya, P. K. R. Maddikunta, M. Parimala et al., "An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture," *Computer Communications*, vol. 160, pp. 139–149, 2020.