Hindawi
Journal of Robotics
Volume 2023, Article ID 9785409, 1 page
https://doi.org/10.1155/2023/9785409

Retraction

Retracted: Research on Clue Mining in Criminal Cases of Smart Phone Trojan Horse under the Background of Information Security

Journal of Robotics

Received 17 October 2023; Accepted 17 October 2023; Published 18 October 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This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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] L. Gang and Y. Wen, "Research on Clue Mining in Criminal Cases of Smart Phone Trojan Horse under the Background of Information Security," *Journal of Robotics*, vol. 2022, Article ID 7568110, 11 pages, 2022.

Hindawi
Journal of Robotics
Volume 2022, Article ID 7568110, 11 pages
https://doi.org/10.1155/2022/7568110

Research Article

Research on Clue Mining in Criminal Cases of Smart Phone Trojan Horse under the Background of Information Security

Li Gang[1] **and Yong Wen**[2]

[1]*Guangxi Police College, Nanning, Guangxi 530028, China*
[2]*School of Artificial Intelligence, Guangxi University for Nationalities, Nanning, Guangxi 530006, China*

Correspondence should be addressed to Yong Wen; 20190304001@stu.gxun.edu.cn

Received 1 December 2021; Accepted 6 January 2022; Published 25 February 2022

Academic Editor: Shan Zhong

With the massive popularity and wide application of Android smartphones, there are more and more malware targeting Android smartphones. Research and analysis Android smart phone Trojan horses can provide corresponding technical support for malware detection on Android smart phones, which has good scientific research significance and broad market value. This work studies and analyzes the existing implantation technology of Android smart phone Trojans and analyzes the basic principles and implementation methods of obtaining root permissions on mobile phones. At the same time, the basic principles and implementation methods of mobile phone Trojan horse hiding are also studied. Through the research of the broadcast receiver model of the Android platform, the background monitoring principle and implementation technology of the mobile phone Trojan horse are analyzed, and the theoretical foundation and technical support are provided for the implementation of the Trojan horse background monitoring program in this article. Aiming at the problem of insufficient training corpus in the event relationship classification task, this work proposes an event relationship classification method based on tritraining. This method first trains three different classifiers based on the labeled dataset. In the collaborative training process, the new labeled event pairs used to expand each classifier are provided by the other two classifiers. For the same unlabeled event pair, the relationship prediction results are consistent; then, the event pair is considered to have a higher classification confidence and is placed in the labeled set of the third classifier after labeling. Finally, a well-trained classifier is used to determine the relationship between the pair of events to be tested by voting. This study constructs a weighted network structure model called the conceptual network and determines its upper weight based on the structural information and text data of the knowledge network. Aiming at the problem of the lack of means for mining-related forms between things, the pheromone strategy of absorbing the ant colony algorithm is proposed, and random walks are performed on the conceptual network. By analyzing the pheromone distribution information in the convergent state, the calculation of the semantic relevance is completed. At the same time, the method of semantic clue discovery is realized. The experimental results show that the human cognitive information contained in the knowledge network can meet the needs of the mining of the related forms of things, and the performance of the semantic correlation calculation method based on the convergence pheromone is close to other random walk methods based on the knowledge network.

1. Introduction

With the development and prosperity of the mobile Internet, smart phones have gradually become an indispensable tool in our daily lives due to their powerful computing power and portability [1]. According to data from the China Internet Network Information Center, there were more than 700 million mobile Internet users in China in 2019, surpassing the number of computer Internet users. Due to the proliferation of smart phones and the privacy of data in smart phones, we also pay more attention to the security of smart phones. Although the open source, open, free, and other features of the Android platform have brought a lot of market share, it also brings many security risks to smartphone users. The makers of smartphone Trojan horses can implant malicious code in normal mobile phone applications and release applications bundled with mobile phone Trojan horses on the Android platform arbitrarily [2, 3]. In 2020, the number of Android smartphone users infected by malware exceeded 10 million, making it the hardest hit area for mobile Trojan horses. The total number of new malware for Android mobile phones in China exceeds 3,000. According to the 2020 security threat report released by the NetQin Security Center, Android smartphones have become the smartphones most concerned about mobile malware.

Network information is closely related to social life, so the security of network information is essential to the normal operation of social life [4, 5]. For individuals, the interference or leakage of network information may cause the infringement of personal privacy information; for enterprises, network information security is related to the normal production and operation of enterprises. Therefore, network information security has become an important issue in social life, and ensuring network information security has become one of the most important tasks of the information supervision department [6]. There are many existing network information security control systems, which can monitor and rectify various network information security issues that may have occurred in the past. However, these methods have different degrees of limitations and deficiencies for today's huge amount of data [7]. The huge amount of big data, faster transmission speed requirements, and more effective value extraction methods make the previous management methods no longer fully suitable for network information security control in the era of big data [8]. Therefore, analyzing the characteristics of big data and establishing a more complete network information security control mechanism based on the characteristics of network information in the big data era is to ensure greater efficiency of big data traffic transmission and realize more secure network information functions. Using the network information security evaluation system to improve the reliability of data security during the entire life cycle of network information data can provide more complete working standards for the high-performance realization of the systematic and scientific development of network security [9].

This work studies the implantation technology of Android smart phone Trojan horses, analyzes the basic methods of mobile phone Trojan horses using the principles of social engineering to achieve deceptive implantation of Trojan horses, and analyzes the basic principles and implementation methods of obtaining mobile phone root permissions. At the same time, it studies the hiding technology of mobile phone Trojans, analyzes the basic principles and implementation methods of file hiding, process hiding, network connection hiding, and kernel module hiding, and analyzes the basic methods of communication hiding technology and collaborative hiding. Then, by analyzing the broadcast receiver model of the Android platform, the background monitoring principle and technology of the mobile phone Trojan horse are studied and analyzed. This work studies the application of the semisupervised learning method based on tritraining in the task of event relationship classification. Aiming at the lack of training corpus in the event relationship classification task, this method uses the tri-training method to train three different classifiers on a small amount of existing manually labeled event relationship datasets and then extracts from the mined ones that contain connectives. In a large number of unlabeled datasets, the training corpus is expanded through simple voting, repeated iterations, and continuous optimization of the classification model, ultimately achieving the goal of improving the performance of event relationship classification. This study tests the execution performance of the semantic relevance calculation model and the semantic clue mining model and designs an experimental program. In the experiment, the calculation results of the semantic relevance calculation model are compared with similar algorithms, and the results show that the model has better calculation accuracy; in addition, the stability of the semantic clue mining model is tested, and the results show that the effective path distribution of the model's feedback is in line with expectations and has basic stability.

2. Related Work

Regarding the current research status of network information security in the context of big data, relevant scholars have made research reports on related issues of big data [10]. The reports all propose that in the era of big data, data security needs to be more effectively protected [11]. It is an important means to maintain information security, and technological advancements play an important role in promoting the progress of information security work. Research believes that in the era of big data, the processing and analysis of information security data have become relatively difficult [12]. In the face of information security challenges under new circumstances, it is necessary to combine multiple methods and means to comprehensively view information security issues. In terms of information security evaluation research, relevant scholars have put forward specific implementation methods for information security evaluation through research and built an enterprise information security evaluation model based on the information security management framework [13].

In the area of network information security strategy research under the background of big data, scholars discussed the necessity of building a national competitive intelligence system based on network information security and proposed measures to build a national competitive intelligence system based on network information security [14]. Researchers proposed that intelligence literacy should be regarded as the core element of information security theory and explored solutions to information security theory in the context of big data [15]. Relevant scholars believe that due to the complexity of network information security, whether it is theoretically or technically, network information security problems cannot be completely solved [16]. Therefore, network information security prevention technologies should be combined with other technologies to use existing technologies. Relevant scholars analyzed the problems facing my country's network information security from the "Prism Gate" incident and put forward the viewpoint of constructing my country's network information security strategy from the perspective of the rule of law [17]. Based on the analysis of the importance of network information security in the era of big data, scholars proposed that network information security must be promoted to the height of national security strategy, unified deployment of the top-level design of network security work, and mobilized the entire society to implement network information security work [18]. Relevant scholars analyzed information security risks that may occur in the context of big data from 9 perspectives, including infrastructure and data processing, and constructed a big data information security risk framework and proposed corresponding solutions [19].

Related scholars use a weakly supervised machine learning method to classify the temporal relationship between events [20]. First, they learn a general classification template from the labeled corpus; then, based on the assumption that "each document contains only one temporal relationship subtype," the documents are clustered according to the temporal relationship subtype, that is, the subtypes of the temporal relationship are consistent. The documents are clustered together.

Related scholars have proposed an extraction method for "event pairs" with causal relationships [21]. The article combines the "relation-oriented template-based" method and the "slot-oriented attribute-based" method to mine "event pairs" resources with sequential relationships. First, it mines the causal "predicate pairs" according to the predefined template; then, it uses the two predefined "type slots" to filter the extracted "predicate pairs" and finally obtains the causal "event pair" resources.

Researchers first search for "concept pairs" with relationships in the text, then use the relationship types defined in WordNet to identify explicit causal "event pairs" in the text, and use the extracted results in the answering system [22]. It uses machine learning methods to classify these two types of event relationships (chronological relationship and causal relationship) on manually labeled corpus with temporal relationship and causal relationship [23]. At the same time, a fully supervised classification method can be used to improve the accuracy of event relationship detection. Related scholars use a fully supervised classifier based on the support vector machines (SVMs) algorithm to perform "relational event pairs" on the four causal relationships (i.e., "causes," "effects," "preconditions," and "postconditions") [24].

The event relationship detection method based on pattern matching can identify more fine-grained event types. Relevant scholars have used pattern matching methods to identify six types of event relationships [25]. Relevant scholars use the method of pattern matching to classify and identify the subtypes of event causality [26]. However, the accuracy of the event relationship detection method that uses pattern matching alone is low. Therefore, researchers propose an event relationship detection method that combines pattern matching and event-oriented element filtering to filter the "relational event pairs" obtained from template matching [27]. Experimental results show that an event relationship detection system that combines pattern matching and event element filtering can effectively improve the accuracy of recognition. Relevant scholars also adopted a fusion of template matching and rule-based methods, and the experimental results show that the performance of the system is significantly better than template matching and rule-based performance [28].

3. Methods

3.1. The Implantation Technology of Android Smart Phone Trojan Horse. Due to the open source and openness of Android, the makers of mobile phone Trojan horses can bundle the Trojan horse programs in some normal applications. When users install these legitimate applications, they also install the Trojan horse programs. Although all Android applications must have digital signatures, in theory, applications with the same digital signature can safely share data resources. Applications with different signatures can also access signature-based APIs by granting permissions to each other. The digital certificate does not require the certification of an authority, so the Trojan horse can complete the signature certification by himself and release the application bundled with the Trojan horse at will to bind a Trojan horse to a legal file or application, and it is necessary to study the source code and vulnerabilities of the application. Therefore, the installation of the Trojan horse also needs to study the vulnerabilities of the application used by the target mobile phone and develop the corresponding binding based on its vulnerabilities. Some Trojan horse programs use popular QR codes to bind malicious codes into legal files (for example, implant the source code of Trojan horses into popular theme files) and use technology to pretend to trick mobile phone users into scanning the QR code. There are also some Trojan horse programs that have been implanted in the mobile phone, using the mobile phone user's contact information and using the user's trustworthy social relationship to mass transit network links containing Trojan horse programs to trick other mobile phone users into opening the link to achieve mobile phone Trojan horses. Therefore, the implantation of Trojan horses requires the use of social engineering methods to target the implantation and spread of Trojan horses.

On the Android platform because root permissions can access all applications and data resources, the makers of mobile phone Trojan horses work hard to study the Trojan horses that obtain root permissions. Although the Android platform only allows the system kernel and very few core programs to run with root permissions, the Android platform allows users or applications to obtain root permissions by modifying the source code of the system kernel or core programs.

Generally, mobile phone users do not have root permissions on Android smartphones, so the makers of mobile Trojan horses can take advantage of vulnerabilities in the Android system to obtain root permissions on mobile phones by flashing the phone, using rootkit technology, and modifying the Android source code.

3.2. Background Monitoring Technology of Android Smartphone Trojan Horse. In the Android system, broadcast is an event generated by the operating system and a mechanism for transferring messages between applications. The broadcast receiver is a component provided for the realization of system broadcasts, which can receive and respond to broadcasts. Intent is the medium used to store broadcast messages. The Android platform can either use Intent to start a component or use send Broadcast() to initiate a system-level event broadcast to transmit messages. The Android system allows programmers to develop a broadcast receiver in the application, and then, register the broadcast receiver to the Android system and notify the operating system that there is such a broadcast receiver waiting to receive the broadcast of the Android system, so that the broadcast receiver can monitor and respond to the broadcast. Each broadcast receiver can receive Intent triggered by one or more events. When an event occurs, the system will send a message to the broadcast receiver.

When the broadcast receiver receives a broadcast message, it creates an extended class that inherits the broadcast receiver and executes the on Receive() function. After executing the on Receive() function, the extended class that inherits the broadcast receiver will be destroyed. If the on Receive() function is not executed within a short period of time, the Android system will consider the program to be unresponsive. Therefore, some time-consuming operations cannot be done in the broadcast receiver. If you need to complete some time-consuming operations, you should send an Intent to the service, and the service will complete it.

As shown in Figure 1, the processing principle for the broadcast receiver is the same regardless of whether it is sending a normal broadcast or an ordered broadcast. The Android system needs to manage the registered Receiver, pass the filtered system message to the Receiver implementation class, trigger the on Receive() function, and pass the Intent object passed by the sender to the on Receive() function. The sender passes the message through the Intent object. Android compares the description in the Intent object with the description in the configuration file Android Manifest.xml, finds out the component (Receiver) that matches the description of the Intent object, and then passes the Intent object as a parameter to the on Receive () in Receiver function. When the broadcast receiver receives the relevant message, it executes the corresponding processing program, such as starting an activity for interaction or opening a service.

In this study, the problem of determining event relations is transformed into a binary classification problem, that is, to classify various event relations separately. The classification results are “this relationship” and “not this relationship.” In the test process, for the test event pair p, the classifiers C1, C2, and C3, respectively, give the classification results R1, R2, and R3. Then, we count the results that appear most in R1, R2, and R3 as the final classification result of p. For example, for binary classification of “causal relationship,” the three classifiers C1, C2, and C3 will give the results of “it is the relationship,” “it is the relationship” and “not the relationship,” and the final relationship of the test event pair is judged as “this relationship,” that is, “causal relationship.”

3.3. Design Ideas and System Framework. Aiming at the problem of insufficient training corpus in the event relationship classification task, this study uses a collaborative training algorithm to propose an event relationship classification method based on tritraining. This method first trains three different classifiers based on the labeled corpus. On this basis, the training set is expanded by selecting samples with higher confidence in the classification from the unlabeled set by majority voting. Then, they use the expanded training set to retrain the classifier, loop and iterate repeatedly, and continuously improve the classification model until the termination condition is met. Finally, three well-trained classifiers are used to determine the event relationship, and the final event relationship type is also output by voting. The system framework of the event relationship classification method based on tritraining is shown in Figure 2.

3.4. Cotraining Method Based on Tritraining. Aiming at the problem of insufficient event relationship training corpus, this study proposes an event relationship classification method based on tritraining. This method uses Gigaword linguistic resources as external data resources, mines event relation pairs containing connectives from it, and uses the relational categories corresponding to connectives as prior knowledge.

On this basis, four features of frame semantics, event trigger word, trigger word part of speech, and event category are extracted to generate an unlabeled event relationship dataset. The labeled set is generated after extracting the same features from a small number of manually labeled event relationship datasets.

In this study, half of the labeled set is used as the training set, and the other half is used as the test set. Then, we use the tritraining method to select higher confidence event pairs from the unlabeled set and add them to the training set and iteratively train the model until the unlabeled set is empty or the size of the unlabeled set no longer changes. Finally, the test set is classified according to the learned model, and the final event relationship classification result is generated.

The tritraining algorithm is designed to repeatedly sample a small number of labeled datasets, train three different classifiers X, Y, Z, and classify and label a large number of unlabeled sample data through the consistency judgment between the three classifiers. In detail, during the training process, the newly labeled samples obtained by any one classifier (for example, classifier X) are jointly determined by the other two classifiers (classifier Y and classifier Z). If two classifiers give the same unlabeled sample x the same classification label L, that is, $Y(x) = Z(x)$, then the unlabeled sample is added to the current classifier X.

It is worth noting that for an unlabeled event relationship pair, when the classification results given by the two classifiers are the same, the same conditions as the prior relationship category must also be met before the event pair will be added. For example, if the prior relationship category of an unlabeled event pair is expansion and the categories given by classifiers C1 and C2 are expansion, then the event pair will be added to the labeled set L, and the category will be marked expansion.

When the cooperative training stop condition is met, that is, when the size of the unlabeled dataset no longer

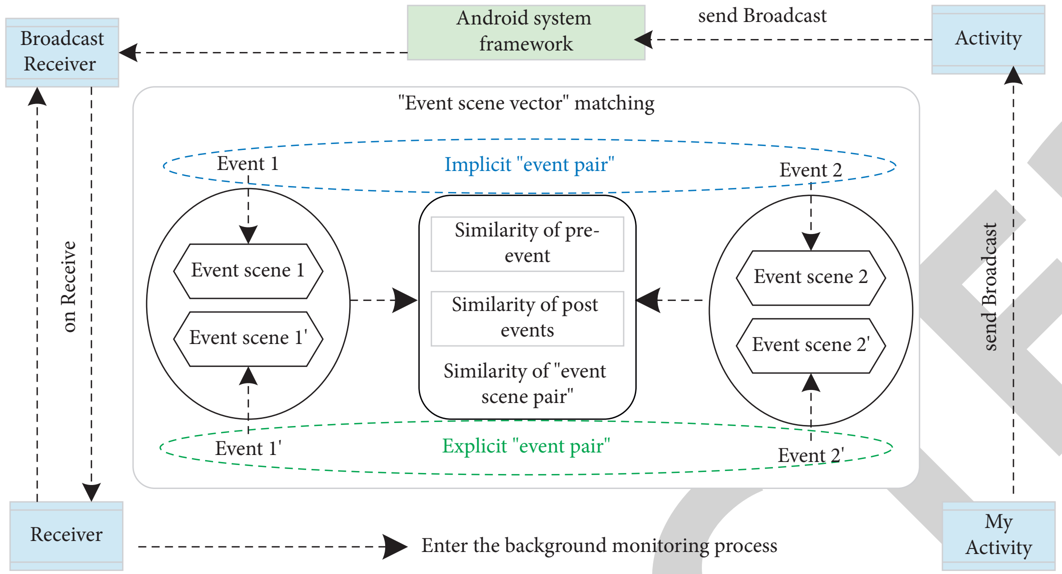


Figure 1: Broadcast receiver model.

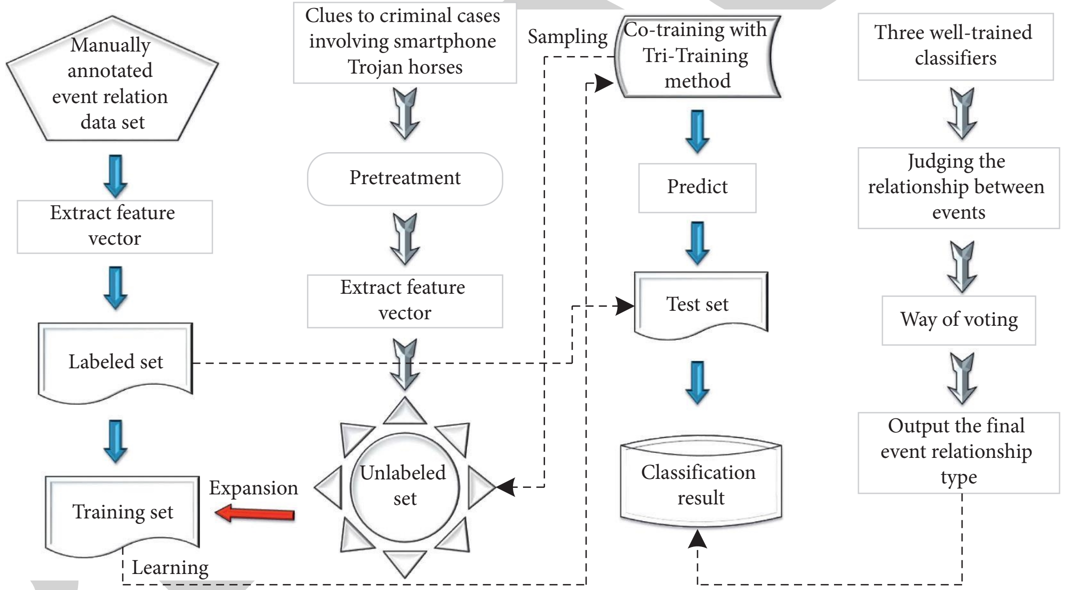

Manually annotated event relation data set
Extract feature vector
Labeled set
Training set
Expansion
Learning
Clues to criminal cases involving smartphone Trojan horses
Pretreatment
Extract feature vector
Unlabeled set
Sampling
Co-training with Tri-Training method
Predict
Test set
Classification result
Three well-trained classifiers
Judging the relationship between events
Way of voting
Output the final event relationship type

Figure 2: Flowchart of event relationship classification method based on tritraining.

changes or is empty, the training is stopped. At this point, three well-trained classifiers C1, C2, and C3 are obtained, and these three classifiers are used to classify and determine the relationship between the event pairs in the test set.

3.5. Implicit Relationship Detection Based on the Conceptual Model. Figure 3 shows the implicit relation detection framework based on the conceptual model of parallel arguments. This study extracts key information from the implicit arguments to be tested, mines parallel arguments containing connection clue words (explicit connectives and functional connectives) based on a large-scale local corpus, and classifies parallel arguments related to the same clue words. The reason is that similar knowledge is involved in similar arguments, which can form semantically targeted conceptual descriptions. For the parallel reference argument set and the implicit argument to be tested after classification, feature extraction and attribute description are performed, respectively, to construct conceptual models A and B (knowledge units formed by unique combination of argument features). The two conceptual models measure parallelism through similarity matching. Through the use of feature vector similarity and conceptual submodel similarity measurement methods, statistical methods are used to obtain conceptual model A with the highest similarity to conceptual model B, and the largest possible text relationship is output through the mapping system as the final result.

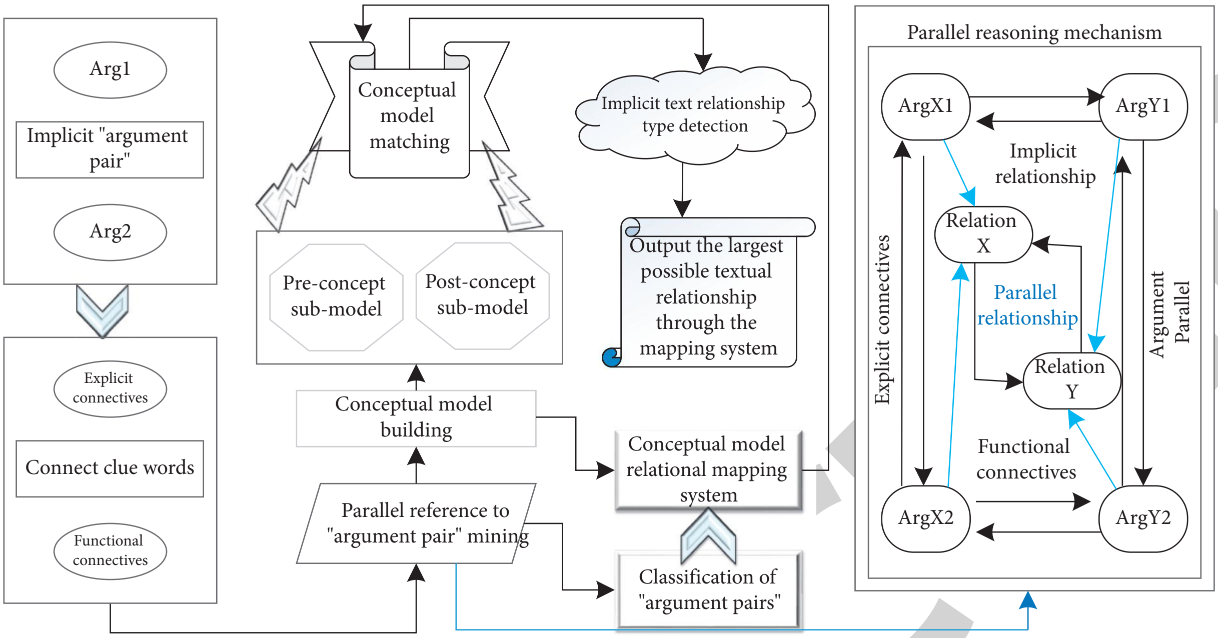

Arg1
Implicit "argument pair"
Arg2
Explicit connectives
Connect clue words
Functional connectives
Conceptual model matching
Implicit text relationship type detection
Pre-concept sub-model
Post-concept sub-model
Output the largest possible textual relationship through the mapping system
Conceptual model building
Conceptual model relational mapping system
Parallel reference to "argument pair" mining
Classification of "argument pairs"
Parallel reasoning mechanism
ArgX1
ArgY1
Implicit relationship
Relation X
Parallel relationship
Relation Y
Explicit connectives
Argument Parallel
Functional connectives
ArgX2
ArgY2

Figure 3: Implicit relationship detection framework based on the parallel argument concept model.

Semantic parallelism calculation consists of two parts: one is the balance calculation model and the other is the parallel calculation model. Among them, the balance calculation model attempts to solve the problem of balance between the preargument Arg1 and the postargument Arg2 in the calculation of semantic parallelism. The parallelism of the preargument Arg1 of the two types of "argument pairs" is PreSim, and the parallelism of the postargument Arg2 is PosSim. After the numerical balance of PreSim and PosSim, the overall "argument pair" is formed. The calculation method is as follows. Among them, the exponential operation of the denominator plays a normalizing role:

$$\mathrm{Sim}\left(a_t, a_c\right) = \frac{\mathrm{PreSim} - \mathrm{PosSim}}{4} e^{-2(\mathrm{PreSim} - \mathrm{PosSim})}. \qquad (1)$$

The parallel calculation model is used to calculate the parallelism PreSim and PosSim. This method uses the vector space model (VSM) as the semantic description structure of the argument and uses the Jaccard algorithm to calculate the parallelism. The Jaccard algorithm is

$$J = \frac{a_t \bullet 2a_c}{a_t \bullet a_c}\left(a_t - a_c\right). \qquad (2)$$

Among them, J is the Jaccard metric value of PreSim and PosSim. a_{t} and a_{c}, respectively, represent the feature vector of Arg1 or Arg2 in the "argument pair" to be tested and the candidate parallel "argument pair." When PosSim is calculated, a_{t} and a_{c} are the feature vectors of Arg2 in the "argument pair" to be tested and the candidate parallel "argument pair," respectively. The numerator is the inner product operation, that is, the weights of the corresponding dimensions of at and ac are, respectively, multiplied and then summed; the absolute value of the denominator part represents the length of the feature vector. The weight of each dimension of the feature vector is calculated by TFIDF, where TF is the frequency of the word feature where the argument is located (obtained after filtering by stop words). The TF of the word features in "pair" is counted, respectively, on the PDTB where the "argument pair" is located and the GIGAWORD local static corpus; the IDF of the antidocument frequency is counted by Gigaword.

An important part of the method framework of this study is the construction of a mapping system that maps conceptual model B to conceptual model A. The mapping of the two conceptual models involves the similarity matching of the two, and the overall similarity measurement method is

$$\mathrm{Similarity} = \frac{\mathrm{PosCSMSim} - \mathrm{PreCSMSim}}{4} e^{-(\mathrm{PosCSMSim} - 2\mathrm{PreCSMSim})}. \qquad (3)$$

Among them, PreCSMSim and PosCSMSim, respectively, represent the similarity between the preconcept submodels and the similarity between the postconcept submodels. The denominator in the formula plays a normalizing role.

The calculation process of PreCSMSim and PosCSMSim involves calculation of the similarity of three pairs of conceptual submodels. Each concept submodel is composed of several clusters, and the elements in each cluster are vectors (such as KeyWordi) that characterize the attributes of this type of cluster. Therefore, the similarity of a pair of preentity concept submodels is the similarity of the two clusters, that is, the final refinement is the similarity of the elements in the clusters. The measurement method is

$$\mathrm{Sim\_}FV(X,Y)=\prod_{k=0}^{N_1} Set_{XY}(k)-\prod_{i=0}^{N_2} OP_{XY}(i)-\prod_{j=0}^{N_3} Val_{XY}(j). \tag{4}$$

Among them, X and Y, respectively, represent the elements in the two clusters (such as KeyWordi and KeyWordj) that need to be similarly calculated, and their similarity is the sum of the similarity weights of the 9-dimensional features. Since each dimension feature has both a numerical form and a word set form, it cannot be directly calculated using the vector space model (VSM).

$$\begin{aligned}\mathrm{Val}_{XY}(j)&=\frac{F_j(X)+F_j(Y)}{\mathrm{Min}\left[S_j(X),S_j(Y)\right]}-\frac{F_j(X)-F_j(Y)}{\mathrm{Max}\left[S_j(X),S_j(Y)\right]},\\ DF&=\log_2 n-\log_2(N-1).\end{aligned} \tag{5}$$

When $j=1$, $F_j(X)$ and $F_j(Y)$ are the position eigenvalues in X and Y; $S_j(X)$ and $S_j(Y)$ are the lengths of arguments to construct X and Y, which are obtained after normalization. When $j=2$, $F_j(X)$ and $F_j(Y)$ are the DF values in X and Y (n is the number of argument categories containing characteristic words, and N is the total number of argument categories); $S_j(X)$ and S_j (Y) are the respective n values. After normalization, the similarity weight of the DF value of X and Y is obtained.

$$\mathrm{Set}_{XY}(k)=1-\frac{G\left[S_k(X),S_k(Y)\right]}{Min\left[N_k(X),N_k(Y)\right]}+\frac{G\left[S_k(X),S_k(Y)\right]}{Max\left[N_k(X),N_k(Y)\right]}. \tag{6}$$

$S_k(X)$ and $S_k(Y)$ represent the single-sentence dependent word set, cross-sentence dependent word set, synonym set, upper word set, and lower word set (according to the value of k) of the respective characteristic words in X and Y; $G(S_k(X), S_k(Y))$ represents the number of word co-occurrences of the feature word set corresponding to X and Y (the number of cross words in the word set); $\mathrm{Max}(N_k(X), N_k(Y))$ represents the respective feature words of X and Y.

The similarity calculation between the preentity concept submodels (consisting of multiple clusters) in the two conceptual models is based on the similarity calculation of feature vectors. The calculation of the similarity between the submodels is the calculation of the similarity between the two clusters. Similarity matching needs to be classified and matched according to the "attribution" of the concept. For example, the pre-subconcept of the "argument pair" to be tested must match the pre-sub-subconcept of the parallel reference "argument pair," but the post-subconcept or the sub-subconcept of the parallel reference "argument pair" cannot be selected. Through the "concept model-text relationship" mapping system, that is, the text relationship to which the connected clue words corresponding to the statistically optimally matched conceptual model belongs, the textual semantic relationship of the "argument pair" to be tested is inferred.

4. Results and Analysis

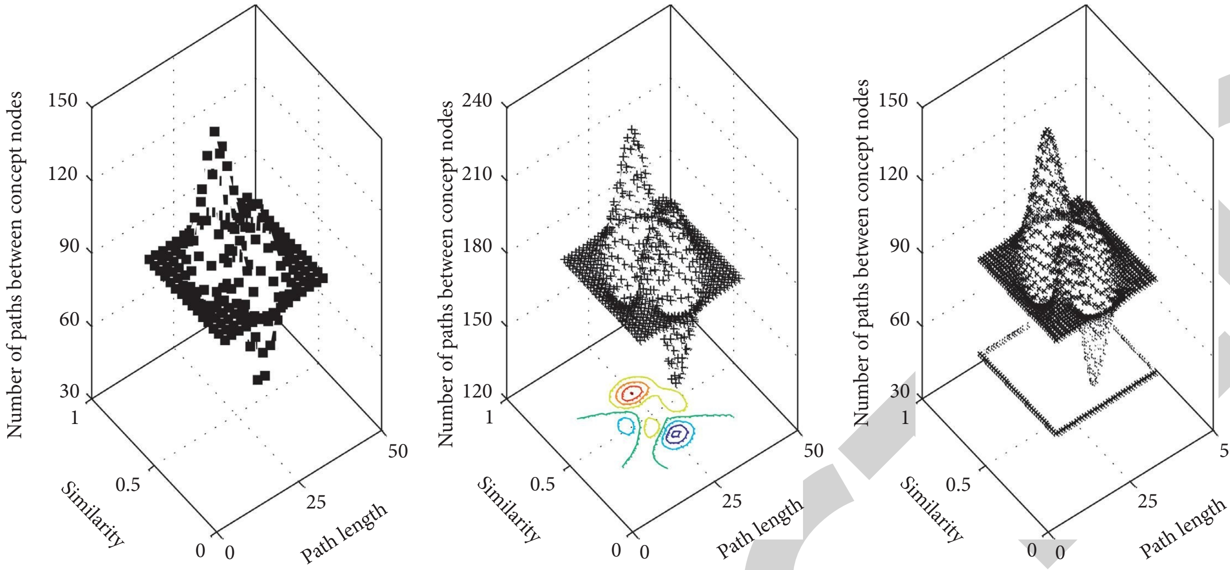
4.1. Model Parameter Setting. In order to determine the influence of different types of subnetworks in the calculation of semantic relevance, three groups of conceptual networks with different components were prepared in the experiment: a conceptual network based entirely on hyperlinks and a concept with 50% similarity between hyperlinks and texts. The network is completely based on the conceptual network of text similarity. Then, on these three groups of conceptual networks, using the data of illegal and criminal cases as the dataset, we perform semantic walks and sort out the path data obtained from the walks to obtain the distribution of the number of paths between conceptual nodes, as shown in Figure 4. The use of the hyperlink structure increases the connectivity between conceptual nodes, increases the effective path between nodes, and reduces the average length of the path at the same time.

The hyperlink structure is added to the conceptual network based on text similarity, and the calculation method of semantic relevance in this study is used to calculate based on the data of illegal and criminal cases. With the difference in the ratio of the hyperlink structure and the text similarity structure, the performance change of the algorithm is shown in Figure 5. It can be seen that when the weight ratio of the concept subnetwork based on text similarity and the concept subnetwork based on hyperlink is close to 0.5, the Pearson correlation coefficient reaches the highest value of 0.78. The following experiment is implemented based on a weight ratio of 0.5.

In this study, the pheromone residual ratio is set to 0.85, and the value of the termination probability is determined through experiments. In order to determine a reasonable value, a test program is set up in this article, and K in TOP-K is set to 50. The length distribution of the obtained path is shown in Figure 6. It can be seen from the figure that most of the paths are paths less than or equal to 240 in length. A small termination probability value will make the wandering agent explore a longer path. Taking into account that the termination probability should be set as far as possible to take into account the path coverage and execution cost, the termination probability is set to 0.3.

4.2. Semantic Relevance Calculation. The semantic relevance calculation model proposed in this study is used to perform semantic calculation on the word pairs, and the calculation results are compared with the average value of the expert evaluation results of the word pairs in the criminal case data, and the Spearman rank correlation coefficient is used as the comparison result measurement index. At the same time, the calculation performance of the algorithm in this study is compared with similar algorithms in the data of crimes.

Figure 7(a) shows the value calculated by the calculation model of semantic relevance in this study for the vocabulary pairs in the data of crimes. Figure 7(b) shows the expert evaluation values. It can be seen from the figure that the two have a certain linear relationship.



(a) (b) (c)

Figure 4: Distribution of the number of paths between conceptual nodes. (a) A conceptual network based entirely on hyperlinks. (b) A conceptual network with 50% similarity between hyperlinks and text. (c) A conceptual network based entirely on text similarity.

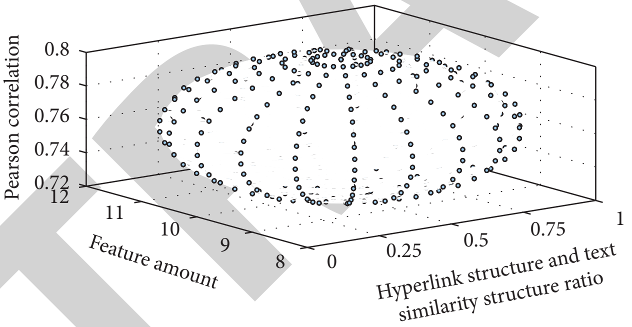


Figure 5: The influence of different ratios of concept subnetworks on algorithm performance (Pearson correlation coefficient).

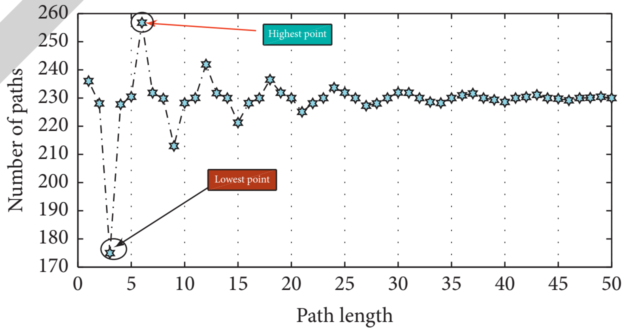
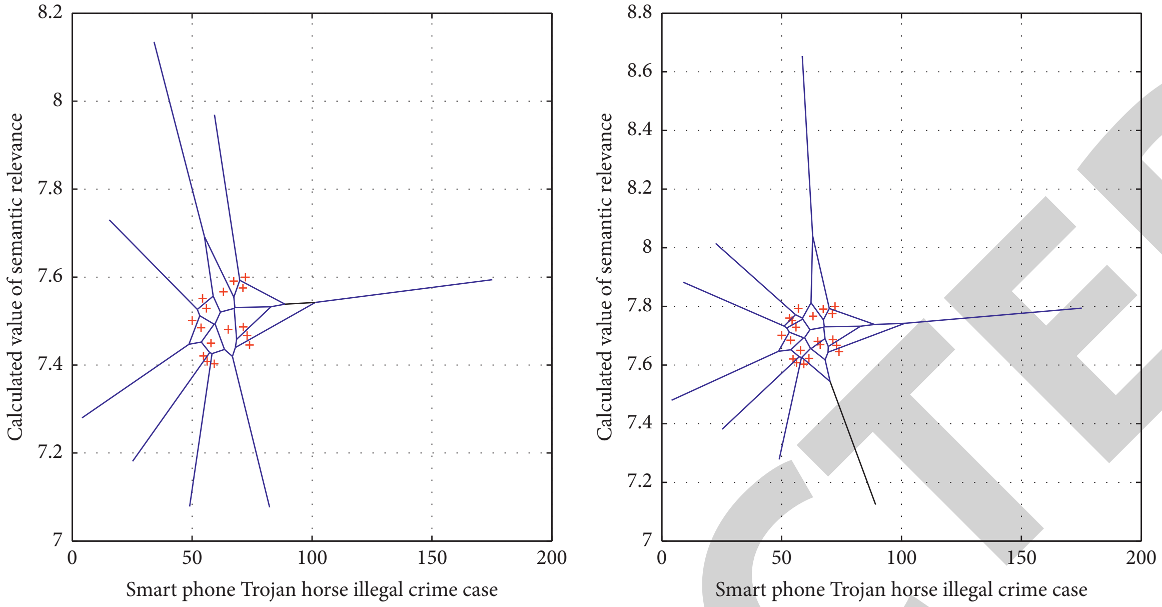


Figure 6: The length distribution of the sampling connection path.

In order to quantify the relationship between the settlement results of the semantic relevance calculation model in this study and the expert judgment standard dataset, to test the calculation performance of the algorithm in this study, the experiment uses Spearman's rank correlation coefficient as a measure, and the correlation coefficient is 0.74.



(a)

(b)

Figure 7: Data on illegal and criminal cases, benchmarks, and algorithmic calculation results. (a) Calculated value of semantic relevance. (b) Expert evaluation value.

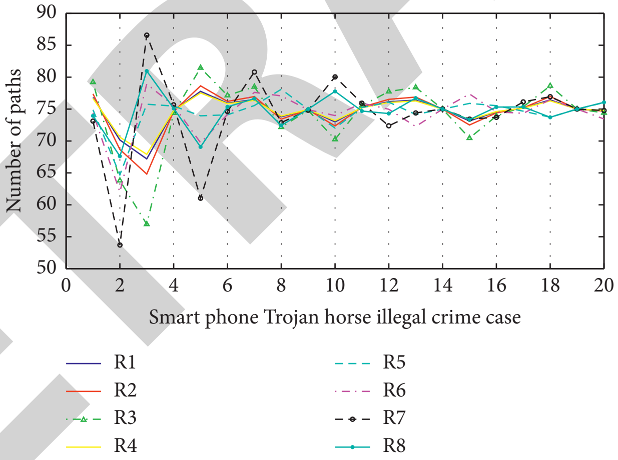


Figure 8: Distribution diagram of the sequence matching of each group of paths.

4.3. Semantic Cue Mining. For the test of semantic clue mining results, this study adopts the method of comparing with the results of manual judgment: this study randomly selects 5 pairs of vocabulary from the data of illegal and criminal cases and divides the experiment into 8 groups according to the vocabulary and denoted by Ri. We map the selected 20 words to the corresponding concept explanation page in the open domain knowledge network and manually click the hyperlink of the page body content from the i1. Special pages encountered during the process are not included in the association path. In the process of exploring the association path of each vocabulary pair, each experiment participant has 5 opportunities to obtain as many different association paths as possible. In this study, Jaccard coefficient is used as a measure of the accuracy of semantic clues.

Figure 8 shows the corresponding distribution of each group of comparison data in TOP-K in the extended example experiment for feedback. Judging from the data of 8 groups of experimental feedback, the experimental data did not appear "inverted," and the experimental results are consistent with the semantic cue criticality ranking returned by the extended algorithm, indicating that the extended feedback data are generally consistent with manual judgment.

In order to verify the reliability and stability of the feedback information provided by the extended algorithm, this work studies and calculates the mean and variance of the

Table 1: Jaccard coefficient mean and coefficient variance.

Group number/serial number	Jaccard coefficient mean					Jaccard coefficient variance				
	1	2	3	4	5	1	2	3	4	5
1	0.62	0.65	0.62	0.73	0.89	5.2	6.4	7.1	6.3	21.3
2	0.67	0.61	0.65	0.77	0.84	5.5	5.2	7.2	6.1	21.4
3	0.63	0.64	0.72	0.76	0.88	5.7	4.7	7.6	6.5	21.2
4	0.65	0.67	0.85	0.82	0.86	5.8	4.5	7.5	6.9	21.7
5	0.66	0.68	0.81	0.74	0.87	5.9	4.9	7.8	6.1	21.8

Jaccard coefficients of the comparison path and the corresponding feedback path. Table 1 provides the calculated mean and calculated variance of top 5. The data show that the minimum average Jaccard coefficient between the algorithm output and the comparison data is 0.61, and the variance is basically stable at a small level.

5. Conclusion

This work studies and analyzes the existing implantation technology of Android smart phone Trojans and analyzes the basic principles and implementation methods of obtaining root permissions on mobile phones. At the same time, it also studied the hiding technology of mobile phone Trojans, analyzed the basic principles and implementation methods of file hiding, process hiding, network connection hiding, and kernel module hiding, and analyzed the communication hiding technology and the basic method of cooperative hiding. Through the research of the broadcast receiver model of the Android platform, the background monitoring principle and implementation technology of the mobile phone Trojan horse are analyzed. Aiming at the problem of insufficient training corpus in the event relationship classification task, a tritraining-based event relationship classification method is proposed. This method trains three different classifiers through the labeled corpus, expands the training set by selecting samples with high classification confidence from the unlabeled set by majority voting, and then uses the new training set to retrain. In the process of conceptual network construction and semantic relevance calculation, "semantic strength" is introduced as a quantitative evaluation index for the contribution and reliability of each semantic element in the semantic information comparison, and corresponding measurement rules are designed. According to the semantic strength of the semantic element, the attention degree of the element obtained in the semantic calculation process is set, so that the extraction and application of the semantic information are more detailed, thereby improving the calculation accuracy of the algorithm. At present, there are not many research studies on semantic clue mining, and the experimental scheme for testing its performance is not mature. Although the mining of semantic clues is similar to the path exploration on the Internet, the integration of semantic elements, the dependence on human intelligence, and the lack of azimuth and distance factors make the general path inspection scheme difficult to apply to the subject of this article. Therefore, this study proposes a quantitative evaluation mechanism for the reliability of semantic clues. Although the author has carried out exploration and research work on this topic, due to the relatively new topic and lack of relevant research and data, the semantic clue mining algorithm of this topic still has major shortcomings and still needs to be improved.

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

The authors acknowledge Project Research and Practice on Teaching Reform of Public Security Based on Mobile Network Security Information Platform supported by special project of 2019, Guangxi Higher Education Undergraduate Teaching Reform Foundation (2019JGB402).

References

[1] J. Espin and S. Perz, "Environmental crimes in extractive activities: explanations for low enforcement effectiveness in the case of illegal gold mining in Madre de Dios, Peru," *The Extractive Industries and Society*, vol. 8, no. 1, pp. 331–339, 2021.

[2] E. Abdelzaher and Á. Tóth, "Defining crime: a multifaceted approach based on lexicographic relevance and distributional semantics," *Argumentum*, vol. 16, pp. 44–63, 2020.

[3] J. Xiong, "A method of mining key accounts from internet pyramid selling data," *Tehnički Vjesnik*, vol. 26, no. 3, pp. 728–735, 2019.

[4] P. Hensinger, "Trojan horse "digital education"–on the road to a conditioning institution set up in a school without teachers," *Current Concerns*, vol. 19, pp. 17–30, 2017.

[5] Y. Wu and J. Zhang, "Building the electronic evidence analysis model based on association rule mining and FP-growth algorithm," *Soft Computing*, vol. 24, no. 11, pp. 7925–7936, 2020.

[6] Y. Li, S. Yang, J. Zheng, Z. Zou, R. Yang, and W. Tan, "Trojan horse DNA nanostructure for personalized theranostics: can it knock on the door of preclinical Practice?" *Langmuir*, vol. 34, no. 49, pp. 15028–15044, 2018.

[7] J. Demers, "Is a trojan horse an empty signifier? the televisual politics of orange is the new black," *Canadian Review of American Studies*, vol. 47, no. 3, pp. 403–422, 2017.

[8] W. Lu, "Big data analytics to identify illegal construction waste dumping: a Hong Kong study," *Resources, Conservation and Recycling*, vol. 141, pp. 264–272, 2019.

[9] L. Xingang, "Analysis of criminal activities exploiting social media: with special regards to criminal cases of WeChat fraud in Chinese jurisdiction," *Journal of Legal Studies "Vasile Goldiş"*, vol. 26, no. 40, pp. 19–36, 2020.

[10] X. Peng, X. Luo, and J. Li, "Meaning construction and judicial identification: difficulties and countermeasures of criminal regulation of illegal fundraising behavior on online P2P lending platforms," *International Journal of Legal Discourse*, vol. 4, no. 1, pp. 47–68, 2019.

[11] M. Wazid, S. Zeadally, and A. K. Das, "Mobile banking: evolution and threats: malware threats and security solutions," *IEEE Consumer Electronics Magazine*, vol. 8, no. 2, pp. 56–60, 2019.

[12] R. Krishnamurthy and E. Chandra, "Development of an efficient attribute weighted fuzzy clustering and optimization using genetic algorithm for crime application," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 18, pp. 1595–1600, 2018.

[13] H. W. Yang, L. C. Huang, and M. S. Hwang, "Research on detection and prevention of mobile device botnet in cloud service systems," *International Journal on Network Security*, vol. 23, no. 3, pp. 371–378, 2021.

[14] Y. H. Chen, H. H. Li, and Z. N. Yu, "Study of repeated e-government project audit based on text mining," *International Journal of Information Technology and Management*, vol. 16, no. 4, pp. 391–404, 2017.

[15] S. K. Mousavi, A. Ghaffari, S. Besharat, and H. Afshari, "Improving the security of internet of things using cryptographic algorithms: a case of smart irrigation systems," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2033–2051, 2021.

[16] O. Frausto and N. L. Castellanos, "State violence, capital accumulation, and globalization of crime: the case of ayotzinapa," *Latin American Perspectives*, vol. 48, no. 1, pp. 202–216, 2021.

[17] H.-J. Zhu, Z.-H. You, Z.-X. Zhu, W.-L. Shi, X. Chen, and L. Cheng, "DroidDet: effective and robust detection of android malware using static analysis along with rotation forest model," *Neurocomputing*, vol. 272, pp. 638–646, 2018.

[18] J. Xiong, M. Tu, and Y. Zhou, "Using weighted similarity to assess risk of illegal fund raising in online P2P lending," *International Journal of Digital Crime and Forensics*, vol. 10, no. 4, pp. 62–79, 2018.

[19] M. P. Masogo and J. T. Mofokeng, "An analysis on illegal online gambling activities: the comparative study within the Gauteng, North West and Limpopo Provinces," *International Journal of Social Sciences and Humanity Studies*, vol. 10, no. 1, pp. 33–48, 2018.

[20] J. Araya, C. Azócar, C. Azócar, and A. Mayol, "Exploring the daily life of mining communities: the case of Antofagasta, Chile," *Rural Society*, vol. 28, no. 3, pp. 226–239, 2019.

[21] G. K. Byemba, "Formalization of artisanal and small-scale mining in eastern Democratic Republic of the Congo: an opportunity for women in the new tin, tantalum, tungsten and gold (3TG) supply chain?" *The Extractive Industries and Society*, vol. 7, no. 2, pp. 420–427, 2020.

[22] K. Marten, "Russia's use of semi-state security forces: the case of the Wagner Group," *Post-soviet Affairs*, vol. 35, no. 3, pp. 181–204, 2019.

[23] F.-C. Tsai, M.-C. Hsu, C.-T. Chen, and D.-Y. Kao, "Exploring drug-related crimes with social network analysis," *Procedia Computer Science*, vol. 159, pp. 1907–1917, 2019.

[24] F. Calvão, "The company oracle: corporate security and diviner-detectives in Angola's diamond mines," *Comparative Studies in Society and History*, vol. 59, no. 3, pp. 574–599, 2017.

[25] G. Crawford and G. Botchwey, "Conflict, collusion and corruption in small-scale gold mining: Chinese miners and the state in Ghana," *Commonwealth & Comparative Politics*, vol. 55, no. 4, pp. 444–470, 2017.

[26] M. Sharma and J. B. Arora, "Cryptography and its Desirable Properties in terms of different algorithm," *IITM Journal of Management and IT*, vol. 8, no. 1, pp. 75–81, 2017.

[27] A. Tripathi, "The economics of hardware trojans: an expert's opinion," *Journal of Information Technology Case and Application Research*, vol. 22, no. 3, pp. 159–174, 2020.

[28] N. Hodgson and S. Ramaekers, "Digitisation, securitisation, and upbringing: interrelations and emerging questions," *Ethics and Education*, vol. 15, no. 4, pp. 391–412, 2020.