

IJACT 20-9-7

The Study of Criminal Lingo Analysis on Cyberspace and Management Used in Artificial Intelligence and Block-chain Technology

¹ Cheolhee Yoon, ^{2*}Bong Gyou Lee

¹ *Researcher, Police Science Institute, Asan., Korea*
E-mail bertter@police.go.kr

² *Professor, Graduate School of Information, Yonsei Univ., Korea*
*E-mail bglee@yonsei.ac.kr*²(Corresponding author)

Abstract

Online cybercrime has various causes. The criminal guilty language, Criminal lingo is active in the shaded area with the bilateral aspect of the word on cyber. It has been continuously producing massive risk factors in cyberspace. Criminals are shared and disseminated online. It has been linked with fake news and aids to suicide that has recently become an issue. Thus the criminal lingo has become a real danger factor on cyber interface. Recently, Criminal lingo is shared and distributed as cyber hazard information. It is transformed that damaging to the youth and ordinary people through the internet and social networks. In order to take action, it is necessary to construct an expert system based on AI to implement a smart management architecture with block-chain technology. In this paper, we study technically a new smart management architecture which uses artificial intelligence based decision algorithm and block-chain tracking technology to prevent the spread of criminal lingo factors in the evolving cyber world. In addition, through the off-line regular patrol program of police units, we proposed the conversion of online regular patrol program for “cyber harem area”.

Keywords: *Criminal lingo, Block-chain, Cyber Harem, Regular Patrol Program*

1. Introduction

The cybercrime occurs by accessing real-time information. At the same time, it exploits the anonymity and proliferation of the Internet. It is being done simultaneously through online sites, online personal media, personal blogs and social networks. Recently, criminal lingo as a cause of cybercrime has continuously produced a risk factor of cyber. Crime language is being shared and disseminated online. As a result, criminal lingo is the basis of cybercrime, fake news, and suicide prevention. Due to the increase of sharing and circulation of factor information through internet and social network, it is easily exposed and spread to factors of cyber harm to youth and ordinary people. Therefore, it is necessary to secure a system for responding to cyber hazard and investigating offenders. In order to prevent it and the proliferation of risk factors in the cyber world gradually, artificial intelligence based decision algorithm is required. At the same time, there is a required to a new smart management architecture design with block-chain technology. For example, if you are investigating an online crime site, the server operated by the offender may be an evidence. However, the actual operating Internet Protocol address changes frequently.

This makes it difficult for the process of collecting evidence to prove a crime. After all, Cyber investigators'

limited computing environment and lack of research personnel power make it difficult to collect and purify criminal gulls. For this reason, it is possible to consider how the artificial intelligence automatically collects and stores cyber shadow areas. The procedure for implementing this is as follows. First, it identifies the crimes that take place in the cyber world. Second, the operating server class is classified by the source IP class and the Internet ISP. Third, we analyze the related cybercrime lingo in real time. Fourth, based on similarity level of regional crime, it identifies period, size and damage range compared with existing cybercrime. Through the above five procedures, it is easy to distinguish cyber shadow areas. If the system is matured and standardized step by step, it may be possible to switch to the online periodic patrol of the police box unit.

In this paper, we use Artificial Intelligence language judgment algorithm for criminal lingo, which is the key to judge cyber harem area detection and judgment.

2. Status Analysis of Cyber Harem

This study is divided into two major areas. First, it is the cybercrime analysis, tracking, and detection of cybercrime lingo. The other is, the part of management through the block-chain. Through this, we have revealed the core of cyber lingo distribution to domestic sites from foreign sites on cyber. The cybercrime environment in Korea and abroad has a transnational approach area and enormous impact power. Cyber-crime information is circulated on the Internet and social networks has become polarized. In addition to, cybercrimes are evolving in new forms such as fake news, suicide aids, online drug trading, hacking, online fraud and SMiShing. In particular, Table 1 shows that factors such as Anonymizer classification of status of cyber harem which we use websites. Among these factors, crime is perceived as threat.

Table 1. Anonymizer classification

Classification	Context
Tor	Through Multi-level node routing to reach the destination and disguised as the final node
Proxy	The client / server relay the communication through itself and conceal the information
HTTP Refer Hide	Initialize system information to be leaked through HTTP Refer Header
Trace back binary filter	Blocking programs running over the web

In order to share their secrets in only a special group, crime data are circulated using a special term. Crime lingo refers to a specific word used in place of a direct mention of an argument for cyber harm, to avoid cyber-cracking of the factor's share and distribution, crackdown. In the case of "first bug", it is often used in social networks such as Facebook, Instagram, and Google in the sense that "if you join the site for the first time, it will charge your cyber money". For example, in Internet illegal gambling, the site operator and the user are using the same language. Korean consonant "ㄱ ㄷ ㄹ ㅅ ㅈ" means "online gambling" or "ladder side". In addition, in the case of "family-park", the park inside the national racetrack was used as a "horse racing gamble online". The use of lingo which is used for cybercrime in domestic and foreign cyber environment constantly being discovered. It is considered that the deep learning technique for the dictionary of the Hanguk Corpus. Cyber sites that distribute and produce criminal lingo are located far away from the influence of the Korean Act. It is located in a place where people or foreigners residing in Korea and overseas are connected and influenced, the server-based online porn sites, suicide aids, and cyber black market websites run in China called the "cyber harem area". They all share the weakness of the management and supervision of online and misuse of the influence of Korean law, and sell personal information DB sales, contract hacking, and drug trade posts. After tracking and detection of criminal lingo analysis, we will use block chains. Related research

has been using block-chain technology through ledger, shared ledger, and distributed ledger.

3. Cyber Criminal lingo detection and collection and analysis technology

3.1. Cyber Crime Lingo Collection

A crime-lingo used for crimes targeting the popular internet media which comprise such services as email, social media sites, blogs and cafes with many users can be tracked and detected. Through Meta Search for large portal search engines such as Google, Bing, Baidu, Naver, Daum, Nate, we tried to detect the criminal language at all times. The main method is a Meta search. The Meta search is a method which a collection server transmits a search query based on a previously prepared corpus (authority dictionary), transmits a keyword to a predetermined search site. In other words, when a search query related to the cyber damage factor information, such as a search query, is transmitted, the Meta search engine automatically classifies a specific website and a specific account. In particular, Figure 1 shows that ways how to get the criminal lingo from the portal website and media collections.

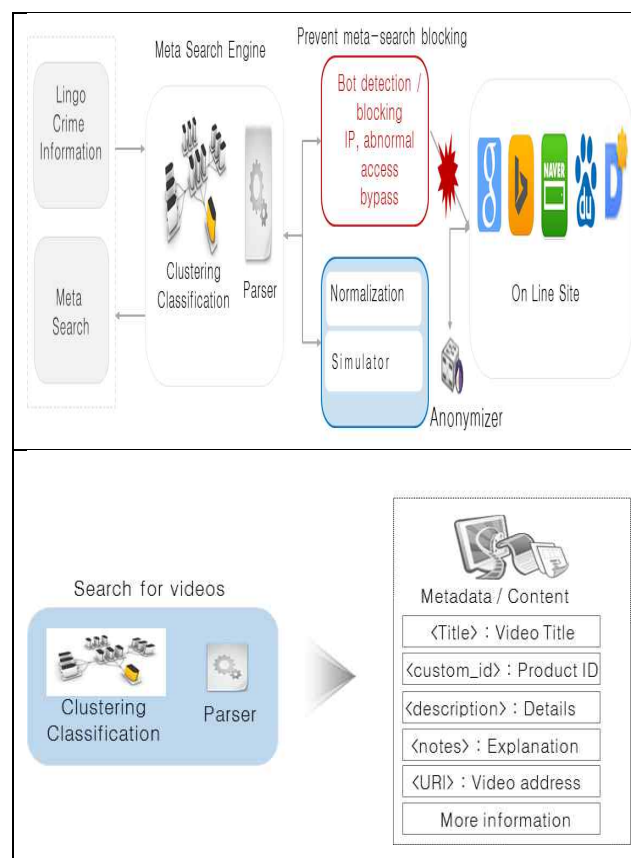


Figure 1. Way of criminal lingo collection

The collected criminals are implemented as a continuous monitoring function for rapid information gathering. A method of disguising as video contents has been used by professional hackers to avoid surveillance. However, since the method and efficiency of inspecting it is not high, it is possible to search the video in the same way as Figure 1. It documented the contents of the YouTube metadata, a shared service.

In addition, the monitoring function of crime guilty lingo as a factor of cyber harm is expressed. The YouTube, we created a data set based on the association analysis for title, publisher, and metadata. Based on the collected videos, it is possible to detect the anomalies and to judge the blocking of criminal words and the relation of crime.

3.2 Cyber Crime Lingo Analysis

AS collections which are likely to be used cybercrime based on the results of the post search results, the statistical value is classified as an important factor. The calculated statistics can be analyzed as follows. First, it is the selection of the importance of the language. Secondly, it is a trend analysis that was exploited for recent crime. Third, statistics on posting frequency and rewriting are calculated based on the data of posting search result. It is possible to handle the analysis of the criminal fluency as shown in Table 2. By analyzing the main channel for the detection of lingo and crime, the selection of lingo could be used in crimes. The resulting statistics can be used to select the importance of fluency and to analyze cyber risk factors and crime trends in recent issues. In particular, Table 2 shows that factors such as identify the main used criminal lingo and the key words by understanding trends [1].

Table 2. Crime lingo analysis process

Classify	The Analysis item
Identify the main used criminal lingo	collected/percentage
Identify the distribution channels	Research on Lingo Distribution
Major lingo that are suspected of being used in actual crime	Proportion of enrolment registration ratio compared investigation
Selected dangerous Criminal lingo	Blocking request crime lingo
Identify the key words by understanding trends	Lingo collection, percentage
Calculate trend by day and time zone	Lingo collection by day, time zone

We used The HanguL morpheme analysis tool in the Konlpy support library (Hwang H., 2015). According to the collected posts, it is categorized as semantic, syntactic lexical. Here, Semantic refers to the analysis of the items that need to create connectivity of collected information by types such as news, blog, twitter, and advertisement. The reason for these linkages must be analyzed through them to collect information such as blogs and advertisements. And Syntactic classifies the names of crime such as ice, stick, and hammer into a syntactic approach for review of the median keyword. Finally, Lexical is an important technical element that categorizes simple collection information such as images, video, and text. It is possible to identify a criminal language and to quickly replace a criminal language, because it is a result of knowing the types of criminals traded online. The analysis's algorithm is used to analyze the sentences of the contents collected by the analysis process of the derivation of the new term. The sentences collected on the bulletin board are separated by morphological analysis of each element and extracted in the form of words. [2]

Figure 1 shows that criminal lingo similarity analysis algorithm such as Sorensen distance, Euclidean distance, Jaccard similarity, nGram, Hamming distance, Levenshtein distance, cosine similarity and soundEx distance metric.

Table 3. Criminal lingo similarity analysis

Algorithm	library	Explanation
Sorensen distance	Distance	Comparison of similarity of two samples
Euclidean distance	Distance	Detection based on string similarity
Jaccard Similarity	SimString	Detecting the degree of intersection between two sets by numerical value
nGram	Ngram	Index search algorithm for index analysis
Hamming distance	Distance	Convert to numbers from the same length
Levenshtein distance	Distance	Determine the similarity of two strings
Cosine similarity	SimString	Document as a number
SoundEx distance metric	Soundex	Judge how similar the two words are

3.3 Apply Block Chain and Smart Contract Application

The Block-chain of publicity is being developed that provides a function to safely and conveniently store data information. Specifically, we use it for “transparency” and “traceability of transaction details” (Kim J., 2016). The Block-chain has established a system in which all transaction details is permanently recorded and managed. In particular, transactions are shared among members of all networks (Kim T., 2017). This principle is used as a crime control book which is included in the distribution of criminal guilt. In this case, it can be used as continuous improvement of information management on cyber harem area in police life safety function. That is, all the risk information collected for the purpose of cyber purge is kept in a block chain structure. Therefore, it is possible to regularly monitor the department managing the cyber-related clean-up business. After that, the information can be stored and managed jointly by the police department and the district unit [3].

In order to create a block containing the Cybercrime Cyber Information, the following technique can be applied. First, the records produced by the information owner are encrypted with the public key. Second, apply a digital signature, a mathematical algorithm. Third, the hash function and the asymmetric key cryptosystem scheme are applied to make anti- forgery and anti- hacking. It is virtually impossible to ensure the stability of the information recorded in the block. Fourth, it transforms it into another text string as a message digest that processed mathematical operations [5].

An encrypted block consists of a block body with a block header and a transaction (No H., 2016). The cyber criminals are created as a block by Meta information. And it provides the retrieved result through the Merkel tree [4]. It extracts the site domain from the inquired post and registers the crime related transaction site keyword as a block hash. After that, through the verification interface, it is classified as “verification success” and “verification failure”, and management becomes possible.

The collected criminal word collecting information can manage by the block-chain [6]. New block hash values such as 000000000000abc09a7a393a8acded75aa67e46cb81f7acaa5ad94f9eacd109, it orders "version", "previous block", "time", "bit", "nonce", "next transaction", "size", "block table of contents", "size", "time received", "relay IP". After that, it is provided as a smart contract as a block chain that can be used by related police officers. It is considered to be a technology that must be used for Block-chain process monitoring, defect analysis, and facility-related data analysis for abnormalities detection [7-10].

4. Conclusion

In this paper, we proposed criminal lingo analysis and block chain technology to detect cyber harem areas. As crime lingo analysis, we use machine learning from word input to result output such as morphological analysis such as Morpheme separation, substantive separation, predicate separation, single morpheme analysis, compound noun and unregistered word processing, Abandon treatment and removal of irradiation

At the base of the dictionary is the constitution of an authority dictionary which is a corpus based on artificial intelligence. The grammatical dictionary, vocabulary morpheme dictionary, field-specific term dictionary and user-defined dictionary should be continuously improved. In addition, classification of a series of processes in which a terminology dictionary for the analysis of criminal fluency is completed is important. The cyber-harmful word corpus was selected as the crime motive and crime type as follows. Money, information damage, publicity, disturbance, reputation, curiosity, retribution, and other types of crimes. Money includes unauthorized use, drugs, gambling, and information corruption, including information manipulation, copyright, and data disclosure. In the subdivision, an authority dictionary containing verbal violence and Internet fraud was created. The analysis process includes morphological analysis of the language and compound noun processing. The re-analyzed result is automatically registered in the corpus dictionary and classification is completed according to the previously learned data. In addition, by applying the block chain, it is stored in the ledger differently from the existing method, and it is continuously improved after distribution to the police department.

In this paper, we tried to block attempts to share criminal data with cyber Lingo crime via online. Then, we tried to track the start of information and collect and manage information through the block chain. Above all based on a professional monitoring system with artificial intelligence, we intend to concentrate investigations and secure work efficiency.

Acknowledgement

This work was supported by Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No.2020-0-00901, Information tracking technology related with cyber crime activity including illegal virtual asset transactions)

References

- [1] H.S. Hwang, "Error Correcting in Korean Morpheme Recovery using Deep Learning", KOREA INFORMATION SCIENCE SOCIETY, Journal of KIISE 42(11), 2015.11, 1454. DOI <https://doi.org/10.5626/JOK.2015.42.11.1452>
- [2] S.Y. Kim, "An Experimental Study on Selecting Association Terms Using Text Mining Techniques", Korea Society for Information Management, Journal of the Korean Society for Information Management 23(3), 2006.9, 160. DOI : <http://dx.doi.org/10.3743/KOSIM.2006.23.3.147>
- [3] J.S. Kim, "A Study on the Factors Affecting the Intention of Acceptance of Block Chain Technology", Soongsil University Ph.D. Thesis, 2016. DOI <https://doi.org/10.2307/249008>
- [4] T.H. Kim, "Block Chain Concept and Case Analysis by Field", Journal of Electrical World Monthly Magazine 487, 2017.7, 58-65(8 pages).
- [5] H.D. Zubaydi, Y.W.Chong, K M.Ko, "A Review on the Role of Blockchain Technology in the Healthcare Domain ". Electronics 2019, 8(6), 679 DOI <https://doi.org/10.3390/electronics8060679>
- [6] Auqib Hamid Lone, Roohie Naaz Mir, "FORENSIC-CHAIN: ETHEREUM BLOCKCHAIN BASED DIGITAL FORENSICS CHAIN OF CUSTODY," Scientific and Practical Cyber Security Journal (SPCSJ) 1(2):21-27, Scientific Cyber Security Association (SCSA), 2017 ISSN: 2587-4667, URL: <https://www.researchgate.net/publication/321746762/download>

- [7] S.H. Kim, S. Chang, and S.W. Lee, "Consumer Trend Platform Development for Combination Analysis of Structured and Unstructured Big Data", *Journal of Digital Convergence*, Vol. 15. No. 6. pp. 133-143, 2017. DOI: <https://doi.org/10.14400/JDC.2017.15.6.133>
- [8] Y. Kang, S. Kim, J. Kim, and S. Lee, "Examining the Impact of Weather Factors on Yield Industry Vitalization on Bigdata Foundation Technique", *Journal of the Korea Entertainment Industry Association*, Vol. 11. No. 4. pp. 329-340, 2017. DOI: <https://doi.org/10.21184/jkeia.2017.06.11.4.329>
- [9] S. Kim, H. Hwang, J. Lee, J. Choi, J. Kang, and S. Lee, "Design of Prevention Method Against Infectious Diseases based on Mobile Big data and Rule to Select Subjects Using Artificial Intelligence Concept", *international Journal of Engineering and Technology*, Vol. 7. No. 3. pp. 174-178, 2018 DOI: <https://doi.org/10.14419/ijet.v7i3.33.18603>
- [10] H.S Jung, J. Kang, C.H. Lee, and S. Lee, "Bigdata Analysis Model for MRO Business Using Artificial Intelligence System Concept", *International Journal of Engineering and Technology*, Vol. 7. No. 3. pp. 134-138, 2018. DOI: <https://doi.org/10.14419/ijet.v7i3.33.18593>