DOI: 10.23977/acss.2025.090214
ISSN 2371-8838 Vol. 9 Num. 2

Research on Network Crime and Security Strategy Based on K-means Cluster Analysis Model

Zimeng Hui[a], Kaiwen Zhao[b,*]

School of Business, Xi'an International Studies University, Xi'an, Shaanxi, China
[a]180913957482163.com, [b]1799366039@qq.com
**Corresponding author*

Keywords: Cyber Security; WCI Index System; Entropy Weight Method; K-Means

Abstract: Cyber security is one of the important issues in global territorial governance, which concerns the security, stability, economic development and public interests of a country and even the whole world. This paper mainly studies the distribution pattern of global cybercrime and establishes the index system of global cybercrime index (GCI). According to the entropy weight method, the top three countries in the global cybercrime index are Indonesia, Tunisia and Nigeria. Countries with an index size above 3.50 are divided according to different geographical characteristics, and the regions with a high proportion of global cybercrime index are Europe, the Pacific region, the tropical region, the Eastern Hemisphere and the coastal region. The K-means cluster analysis model is established, and it is concluded that the countries with high density of cyber crimes include Indonesia, Tunisia, Nigeria, etc. Countries with high success rates include the United States, Switzerland, Serbia, etc. Countries with high rates of reported cybercrime incidents include Albania, Argentina and Armenia. Countries with high litigation rates include Panama, South Korea and Lithuania. The global distribution of cybercrime presents a relatively common pattern, which requires countries to prevent and improve laws and policies in different regions.

1. Introduction

In the current wave of digitalization, whether it is the production and operation of enterprises, or the daily life of people, we are increasingly dependent on the network, and the world is becoming more connected. However, at the same time, cyber crimes are becoming increasingly rampant, and cyber security incidents seriously threaten network security and the normal operation order of economy and society. As shown in Figure 1, it presents the global distribution of cybercrime.

Cyber security is a complex transnational issue, and many cyber security incidents are difficult to respond to due to their cross-border nature, and many organizations would rather pay a fee than let their customers know that they have experienced a security breach. Based on this, many countries have developed their own cybersecurity policies. The International Telecommunication Union (ITU), as the United Nations agency responsible for information and communication technologies (ICT), continues to develop international standards and assessment methods related to cybersecurity, and timely identify and deploy relevant risk mitigation measures to address the growing risks and costs of cybersecurity. Network is one of the important topics in global territorial governance [1], and whether it can effectively deal with various types of network security threats is directly related to the security and stability, economic development and the realization of public interests of a country and even the world. Therefore, it is of great strategic significance to construct relevant network security models.

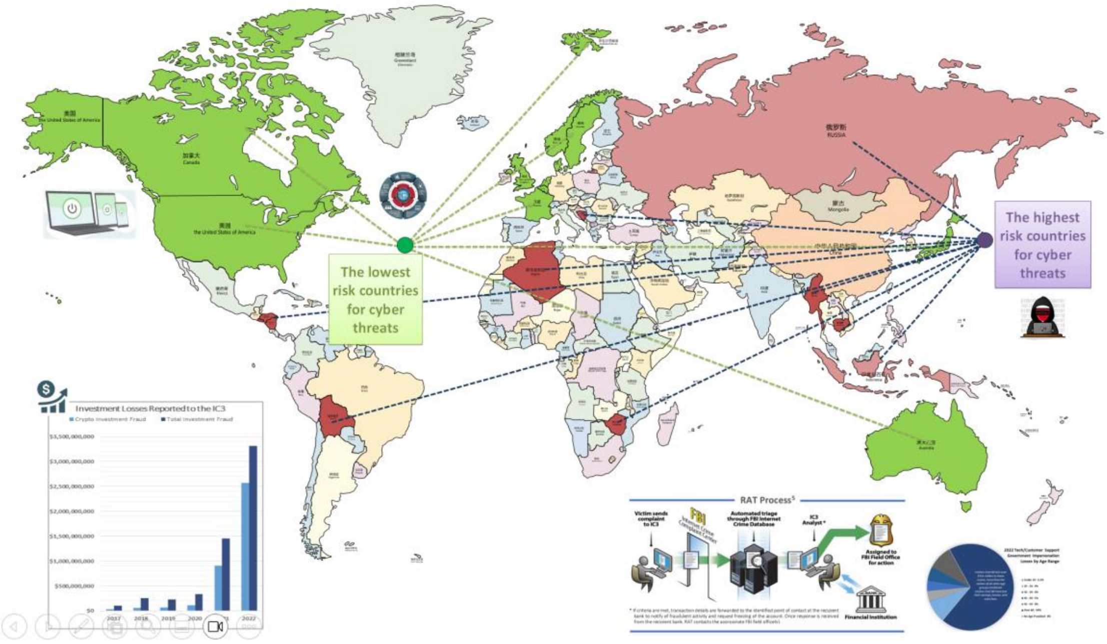
The lowest risk countries for cyber threats
The highest risk countries for cyber threats
Investment Losses Reported to the IC3
Crypto Investment Fraud
Total Investment Fraud
$3,500,000,000
$3,000,000,000
$2,500,000,000
$2,000,000,000
$1,500,000,000
$1,000,000,000
$500,000,000
$0
2017
2018
2019
2020
2022
RAT Process
Victim sends complaint to IC3
FBI Internet Crime Complaint Center
Automated triage through FBI Internet Crime Database
IC3 Analyst *
Assigned to FBI Field Office for action
Financial Institution
2022 Tech/Customer Support Government Impersonation Losses by Age Range

Figure 1: The distribution of global cybersecurity incidents in recent years

To explore what national cybersecurity policies and laws can be designed to be "data-driven," our team will first collect data and reports to build a global cybercrime indicator system, and second, we will calculate the size of the national cybercrime index using the entropy weight method. Then, we build a K-means cluster analysis model to analyze how network security incidents are distributed globally, why such places become the most frequent sites of network security incidents, where more network security incidents will be successful, where more will be thwarted or blocked, and where will proactively report network security incidents. Where these events are handled. Finally, we explore countries' published national security policies and compare them to the distribution of cybercrime, analyzing which policies or legal models can effectively curb cybersecurity incidents and which policies and recommendations can help reduce the level of cybercrime globally.

2. GCI distribution exploration model based on entropy weight method and K-means model

2.1. Data Description

Due to the variety of forms of cybercrime and the scarcity of digital evidence, researchers face many difficulties and challenges in the statistics and reporting of related cybercrime incidents[2]. Therefore, considering the rapid growth of cybercrime in recent years, in order to better measure and comprehensively analyze cybercrime and to ensure the accuracy and scientific nature of the model built, this study selects the period from 2020 to 2023 as the time range for investigation. Data were collected, screened, and cleaned from sources such as the VERIS Community's Json database, SEON, the International Telecommunication Union (ITU)[3], the UNODC Global Crime Database, the US Internet Crime Complaint Center, the Global Email Threat Report, and the World Internet Development Report. A total of 120 countries and 6,918 samples were included in the initial dataset.

First, 33 invalid country samples and 645 invalid samples were removed. Then, missing values were filled using interpolation to ensure data completeness. Next, outliers in the samples were removed. Finally, the data were standardized to complete the processing of the overall data sample.

2.2. Establishment of the Index System of Global Cybercrime Index (GCI)

To make the measurement of cybercrime more universal and credible, this paper first establishes the indicator system of the Global Cybercrime Index from three dimensions: cybersecurity, website risk, and damage loss, as table 1 follows:

Table 1: Index System of Global Cybercrime Index

Dimension	Measurement indicators	Unit	Properties	Weights(%)
Network security	World Cybersecurity Index (WCI)	/	+	20.15
	Comprehensive Cybersecurity Score (CSI)	/	+	15.23
Website risk	Cyber Exposure Index (CEI)	/	-	16.06
	Number of Phishing Email Attacks as a Percentage (NPEA)	%	-	13.07
Damage from hazards	Number of cybercrime incidents (NCI)	ten thousand pieces	-	12.99
	Amount of cybercrime losses (CLA)	million dollars	-	22.50

The cybercrime index of each country in the world in 2023 is calculated by the entropy weight method to visualize and analyze the distribution of cybercrime globally. Next, metrics were measured for different target regions, where different types of regions may have different or intersecting cybercrime outcomes. This is shown in the table 2 below:

Table 2: Different Target Regions

Study area	Metrics	Calculation of indicators
High target area	Number of cybercrime incidents (NCI)	/
Victory zone	Cybercrime Success Rate (CSR)	Number of statistically significant cybercrime payments recorded/number of cybercrime incidents occurring
Frustration zone	Cybercrime Success Rate (CLR)	1 - Cybercrime success rate
Reporting area	Cybercrime Reporting Rate (CRR)	Number of national reports of cybercrime/incidents of cybercrime
Prosecution area	Cybercrime Prosecution Rate (CPR)	Number of statistically actionable cybercrime prosecutions/incidents of cybercrime

2.3. K-means Clustering Model

K-means clustering analysis, also known as K-means clustering[4], is a distance-based clustering algorithm. The goal of this algorithm is to partition the dataset into k clusters through iterative optimization, minimizing the sum of squared distances between data points and their respective cluster centers. The K-means algorithm requires the initial number of clusters k and the initial cluster centers to be specified in advance. It then iteratively updates the cluster centers based on the similarity between data features and the cluster centers, reducing the within-cluster sum of squared errors. The clustering process terminates when the objective function converges or the within-cluster sum of squared errors (SSE) no longer changes, yielding the final clustering results.

The analysis of cybercrime distribution is conducted by collecting data from nearly 100 countries on the number of cybercrime incidents, the success rate of cybercrimes, the reporting rate of cybercrimes, and the prosecution rate of cybercrimes. Among these, the number of cybercrime incidents, the success rate of cybercrimes, the reporting rate of cybercrimes, and the prosecution rate of cybercrimes are the four features in the dataset, while countries are the data points. The frequency of cybercrime occurrence is divided into four levels: high-density crime areas, second-high-density crime areas, second-low-density crime areas, and low-density crime areas, which correspond to the four groups in the K-means clustering.

Firstly, we select the initial cluster centers. To obtain better initial cluster centers, we employ the K-means++ method here[5]. Assuming we want to divide countries into four groups (k=4), we need four initial cluster centers. We randomly select the four feature values of one country as the first initial cluster center. Then, calculate the distance from each country to the selected initial cluster center and choose the four feature values of the country that is farthest from the initial cluster center as the second initial cluster center. This process is repeated until k initial cluster centers are selected. Then let the dataset be $X = \{x_1, x_2, \ldots, x_n\}$, where $x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})$ represents the four feature values of the i-th country; the initial cluster centers are $C = \{c_1, c_2, \ldots, c_k\}$, where $c_j = (c_{j1}, c_{j2}, c_{j3}, c_{j4})$ represents the four feature values of the jth initial cluster center. The Euclidean distance to each initial cluster center $\mathrm{C_j}$, using the following formula.

$$d(x, c_i) = \sqrt{\sum_{j=1}^{d} (x_j - c_{ij})^2} \tag{1}$$

x_j and C_{ij} are the values of x and C_i in the j-th dimension, respectively; d is the dimensionality of the data object; x is the data point; and C_i is the i-th cluster center. The objective function of K-Means is:

$$J = min \sum_{i=1}^{m} \sum_{k=1}^{k} w_{ik} ||x_i - c_k||^2 \tag{2}$$

w_{ik} is the assignment variable. If the data point belongs to the cluster, then w_{ik}=1; otherwise, w_{ik}=0. Substituting this in, we get:

$$d_{ij} = \sqrt{(x_{i1} - c_{j1})^2 + (x_{i2} - c_{j2})^2 + (x_{i3} - c_{j3})^2 + (x_{i4} - c_{j4})^2} \tag{3}$$

We assign each country xi to the cluster corresponding to the initial cluster center C_j that is closest to it. For each cluster, recalculate its cluster center to ensure that during the update process, the cluster center moves toward the centroid of the data points within the cluster. The formula for updating the cluster center is the mean of all data points within the cluster, as follows:

$$c_{j1}^{new} = \frac{1}{|S_j|} \sum_{x_i \in S_j} x_{i1} \tag{4}$$

$$c_{j2}^{new} = \frac{1}{|S_j|}\sum_{x_i \in S_j} x_{i2} \quad (5)$$

$$c_{j3}^{new} = \frac{1}{|S_j|}\sum_{x_i \in S_j} x_{i3} \quad (6)$$

$$c_{j3}^{new} = \frac{1}{|S_j|}\sum_{x_i \in S_j} x_{i3} \quad (7)$$

$$c_{j4}^{new} = \frac{1}{|S_j|}\sum_{x_i \in S_j} x_{i4} \quad (8)$$

Here S_j is the set of data points in the j-th cluster, and $|S_j|$ is the number of data points in cluster S_j.Repeat steps (2) and (3) until the positions of the cluster centers no longer change significantly or the maximum number of iterations is reached.

3. GCI Study and Distribution Results

3.1. Global Cybercrime Index Result

First of all, we use GCI, a more comprehensive and comprehensive score, as a reference for the current situation of cybercrimes in countries around the world, and conduct an overall ranking of GCI. Researchers find that the higher the GCI, the more complex and serious the occurrence of cybercrimes, and they select the top 30 countries in GCI, as shown in the following table 3:

Table 3: Top 30 countries in the global cybercrime index

Rank	Country	GCI	Rank	Country	GCI	Rank	Country	GCI
1	Indonesia	0.724	11	Bangladesh	0.677	21	Pakistan	0.645
2	Tunisia	0.718	12	Peru	0.669	22	Kyrgyzstan	0.638
3	Nigeria	0.717	13	Maldives	0.667	23	Korea	0.628
4	Britain	0.716	14	Laos	0.667	24	Myanmar	0.615
5	Philippines	0.707	15	Uzbekistan	0.660	25	Zimbabwe	0.612
6	Uganda	0.690	16	Zambia	0.659	26	Nepal	0.609
7	Brazil	0.688	17	Cameroon	0.653	27	USA	0.590
8	Angola	0.686	18	Russia	0.649	28	Cambodia	0.583
9	Morocco	0.683	19	Armenia	0.648	29	Oman	0.580
10	Tanzania	0.681	20	Sri Lanka	0.645	30	Namibia	0.578

According to the above table, we can intuitively see that the top three countries in the global cybercrime index are Indonesia, Tunisia and Nigeria, with cybercrime indexes of 0.724, 0.718 and 0.717, respectively, which indicates that for Southeast Asia and Africa and other countries that are still in the stage of development, there is greater room for development and difficulties in the areas of cybersecurity awareness, cyber-risk prevention capacity[6], and the promulgation of laws and policies, and national governance. This indicates that for countries in Southeast Asia and Africa, which are still in the development stage, there is more space and difficulties in the development of cybersecurity awareness, cyber risk prevention capability, enactment of legal policies and national governance.

After sorting the overall GCI we also selected 65 countries with an index score of 3.50 or above to represent the coordinates where cybercrime mainly occurs in the world, and divided these countries according to different geographic features, from left to right, and from the outside to the inside, respectively, according to the continents, oceans, temperature zones, the eastern and western hemispheres, and the coasts and the inland for the division of the five categories in order to better explore the distribution of cybercrime around the globe.

3.2. K-means model clustering results

First of all, the ANOVA of each variable brought into the cluster analysis model obtained that each indicator is significant at the 1% level, indicating that there is a significant difference between the categories we classified in the cluster analysis, which can be brought into the model for the output of the clustering results. Then, according to the classification of the clustered categories, the results of the frequency statistics and dimensionality reduction, the visualization of the distribution of the four categories under each different target region is shown in the following figure 2 and figure 3:

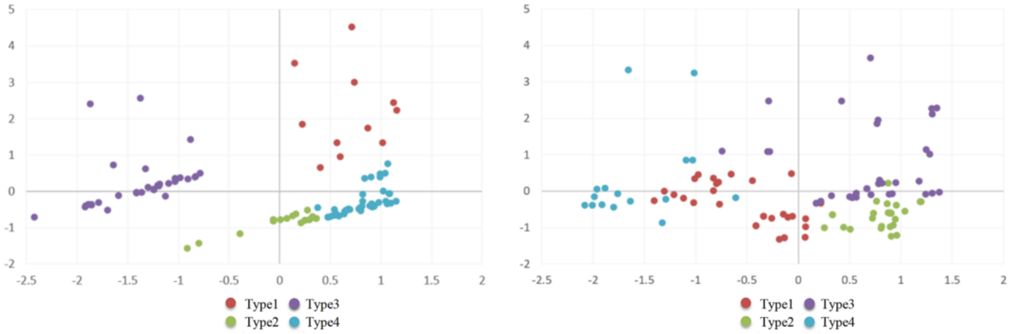

Figure 2: K-means scatterplot of cybercrime high target and success areas

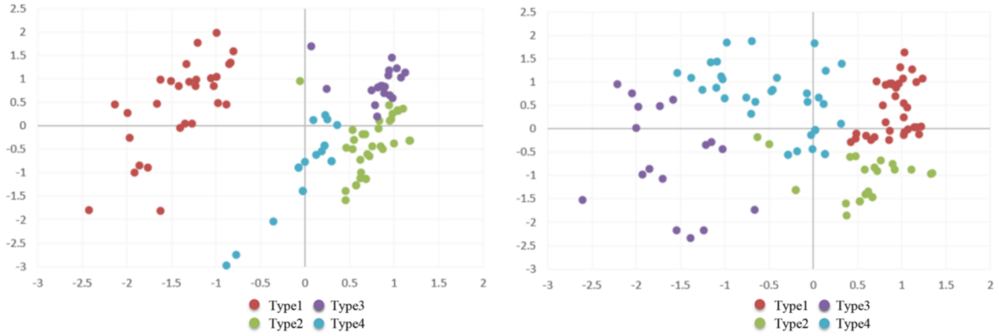

Figure 3: K-means scatter plot of cybercrime reporting and prosecution areas

We added the number of countable global cybercrime incidents, the measured cybercrime success rate, the cybercrime reporting rate, and the cybercrime litigation rate to the indicators according to the requirements of the target region, and after cluster analysis and hierarchical categorization using SPSS and Python, we classified the frequency of occurrence of the representative indicators of the target region of the observed countries into four categories from high to low. Defining category 1 as the highest-ranking country, category 2 as higher, category 3 as lower and category 4 as the lowest-ranking country, the top 10 countries were selected based on the combined ranking of the indicators in category 1 as the representative countries for the high-targeted cybercrime areas, successful areas, frustrated areas, reporting areas and litigation areas of the topic, respectively, as shown in the figure 4 below.

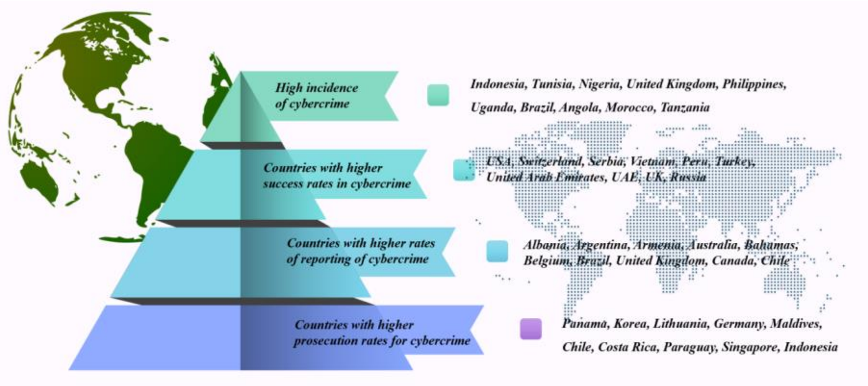
High incidence of cybercrime
Indonesia, Tunisia, Nigeria, United Kingdom, Philippines, Uganda, Brazil, Angola, Morocco, Tanzania
Countries with higher success rates in cybercrime
USA, Switzerland, Serbia, Vietnam, Peru, Turkey, United Arab Emirates, UAE, UK, Russia
Countries with higher rates of reporting of cybercrime
Albania, Argentina, Armenia, Australia, Bahamas, Belgium, Brazil, United Kingdom, Canada, Chile
Countries with higher prosecution rates for cybercrime
Panama, Korea, Lithuania, Germany, Maldives, Chile, Costa Rica, Paraguay, Singapore, Indonesia

Figure 4: Four Categories of Cybercrime Incidents are Representative of Countries

4. Conclusions

By constructing the Global Cybercrime Index indicator system, we quantify the cybercrime outcomes across various target regions, derive measurement indicators, and rank the results obtained through this indicator system. Notably, Indonesia, Tunisia, and Nigeria emerge as the top three countries in the global cybercrime index. To delve deeper into the global distribution of cybercrime, countries scoring 3.50 or above on the index were categorized based on distinct geographical attributes. The regions that contribute the highest proportions to the global cybercrime index include: Europe, Asia, and Africa; the Pacific and Atlantic Oceans; tropical areas; the Eastern Hemisphere; and coastal regions. Furthermore, a K-means cluster analysis model was established to classify countries based on their cybercrime characteristics. After comprehensive ranking, we identify: Countries with high cybercrime density, such as Indonesia; Countries with high cybercrime success rates, exemplified by the United States; Countries with elevated cybercrime incident reporting rates, like Albania; Countries with high litigation rates due to cybercrime, including Panama.

Ultimately, the global distribution of cybercrime reveals a relatively dispersed pattern, closely intertwined with the five dimensions of the ITU's Global Cybersecurity Index (GCI). We please note that while the content is similar in structure and detail, some slight variations in phrasing and terminology have been introduced to maintain readability and clarity.

Here are two advice for the future development to deal with the global cybersecurity. Countries need to ensure that cybersecurity policies are strictly enforced and effectively monitored to address emerging threats and raise cybersecurity standards. In addition, it can work with other countries and organizations to develop and implement cybersecurity policies and share information and best practices to address global cyber threats.

References

[1] Tan Youzhi. Global Governance of Cyberspace: International Situation and Chinese Path[J]. World Economy and Politics, 2013(12):18.

[2] Kwon D, Borrion H, Wortley R.Measuring Cybercrime in Calls for Police Service[J].Asian Journal of Criminology, 2024,19(3):329-351.

[3] Hegarty K,Wilken R,Meese J, et al.Shaping infrastructural futures: The International Telecommunication Union's visions for mobile communications and the anticipatory politics of 5G standardization[J].Mobile Media & Communication, 2025,13(1):171-191.

[4] Xiang W, Yuanhao M,Hao L, et al.Clustering Optimized Portrait Matting Algorithm Based on Improved Sparrow Algorithm[J].Tehnički vjesnik,2023,30(6):1911-1919.

[5] Parveen S, Yang S M. Lasso-Based k-Means++ Clustering [J]. Electronics, 2025, 14(7): 1429-1429.

[6] Kalpit S,Arunabha M. Kernel na ïve Bayes classifier-based cyber-risk assessment and mitigation framework for online gaming platforms[J].Journal of Organizational Computing and Electronic Commerce,2021,31(4):343-363.