Cybercrime Distribution and Policy Assessment Based on K-means Clustering and AHP Model

Jiabao Guo[1,a,*], Xiang Gu[2,b], Yidi Yu[1,c], Mandi Wu[3,d], Mengfei Liu[4,e], Yue Liu[1,f], Huang Huang[5,g]

[1]School of Economics, Beijing Technology and Business University, Fucheng Road, Haidian District, Beijing, China

[2]School of Computer and Artificial Intelligence, Beijing Technology and Business University, Fucheng Road, Haidian District, Beijing, China

[3]Digital Academy, Beijing Technology and Business University, Fucheng Road, Haidian District, Beijing, China

[4]School of Light Industry Science and Engineering, Beijing Technology and Business University, Fucheng Road, Haidian District, Beijing, China

[5]School of Mechanical and Power Engineering, Nanjing Tech University, Puzhu Road, Nanjing, Jiangsu Province, China

[a]3165860945@qq.com, [b]guxiang20060508@sina.com, [c]yuyidi2022@163.com, [d]wumandirdfz@126.com, [e]mengfei0616lmf@163.com, [f]3549154137@qq.com, [g]1793967836@qq.com

*Corresponding author

Keywords: Cybercrime, Cybersecurity Policy, K-Means Clustering, Analytic Hierarchy Process, Global Distribution

Abstract: This study investigates the global distribution of cybercrime and the effectiveness of cybersecurity policies using a comprehensive analytical framework. Cybercrime has become a significant global challenge, causing substantial economic losses and threatening information security worldwide. To address this issue, we employ the K-means clustering method to classify cybercrime data and reveal its geographical distribution patterns. Additionally, the Analytic Hierarchy Process (AHP) is utilized to construct the National Cybersecurity Index (NCI), which evaluates the effectiveness of cybersecurity policies in different countries. Our findings indicate that certain regions are more vulnerable to cybercrime due to factors such as economic conditions, internet penetration rates, and enforcement capabilities. Moreover, the study highlights the importance of policy quality and implementation in enhancing cybersecurity. The results provide valuable insights for policymakers to optimize cybersecurity strategies and improve policy effectiveness. This research contributes to the field by offering a systematic approach to understanding the complex relationship between cybercrime distribution and cybersecurity policies on a global scale.

1. Introduction

As digital technologies expand and interconnectivity deepens, cybercriminals increasingly exploit vulnerabilities for financial, political, and malicious purposes [1]. Cybercrime has evolved from isolated incidents into large-scale, high-intensity attacks that affect multiple nations simultaneously, often intersecting with geopolitical tensions, economic disruptions, and institutional weaknesses. This interconnectedness suggests that cyber threats do not operate in isolation but rather form part of a broader, dynamic risk landscape [2].

Beyond immediate consequences, cybercrime exhibits a cascading effect, akin to the butterfly effect, where a single attack can trigger a chain reaction of related criminal activities and systemic disruptions [3]. The complex interdependencies within the digital ecosystem mean that even minor breaches can escalate into widespread crises, affecting critical infrastructure, financial systems, and national security [4]. Combating cybercrime is increasingly challenging due to its transnational

  DOI: 10.25236/icmmct.2025.015

nature,evolving offender tactics, and limitations in law enforcement. Robust national cybersecurity policies are essential to mitigating these threats [5, 6].

This study aims to construct a comprehensive analytical framework to systematically assess the distribution characteristics of global cybercrime and the effectiveness of cybersecurity policies in different countries [7]. Specifically, this study will employ the K-means clustering method to classify global cybercrime data and reveal the geographical patterns of cybercrime [8]. Meanwhile, the Analytic Hierarchy Process (AHP) will be utilized to construct the National Cybersecurity Index (NCI)to evaluate the effectiveness of cybersecurity policies [9, 10, 11]. Additionally, a game-theoretic model will be applied to analyze the interactive behaviors between national policies and cybercrime groups [12, 13]. Through these methods, this study will not only provide a global perspective on cybercrime distribution but also offer scientific decision-making support for policymakers to optimize the formulation and implementation of cybersecurity policies.

2. Dataset

The data of this study mainly come from official cybersecurity reports released by various governments, crime statistics yearbooks, and the Global Cybersecurity Index (GCI) of the International Telecommunication Union (ITU). At the same time, it combines the cybercrime reporting data provided by international institutions such as the VERIS Community Database (VCDB), the FBI's Internet Crime Complaint Center (IC3), and Europol to obtain detailed information on cyberattacks in different countries. To further enrich the macro variables, this study integrates the Guotai National Accounting Database and the CNRDS Database, collects economic, social, and policy data, and combines existing academic literature to obtain qualitative analysis of the development of cybercrime in different regions. All data have undergone strict data cleaning, including the removal of missing values, interpolation and completion, outlier detection and elimination, and standardization and Winsorization tail-trimming processing to ensure data integrity and reliability and to reduce the influence of extreme values. Statistical analysis uses the R language and Python, and employs tools such as ggplot2 and dplyr for data visualization.

3. Global Distribution Patterns of Cybercrime

By revealing the spatial distribution patterns of cybercrimes and the current situation of their crackdown, we can identify the factors influencing the distribution and prevention of cybercrimes. In this section, analyzing the occurrence frequency and geographical distribution of cybercrimes globally can help identify which countries or regions are most likely to become targets of cybercrimes, thereby facilitating an understanding of the geographical characteristics of the crimes. Further verification of the explanations provided earlier through the current situation of cybercrime crackdowns can uncover more lessons to be learned or warnings. Identifying the key factors influencing the distribution and effectiveness of the prevention of cybercrimes, such as technological, legal, and socio-economic factors, can help us understand why certain countries are more vulnerable to attacks while others have stronger prevention capabilities.

3.1. Global Cybercrime Index and Risk Index

3.1.1. Global Cybercrime Index

This study refers to the research conducted by Miranda Bruced et al. Firstly, the types of cybercrimes are classified into five categories as shown in Figure 1.

For each type of cybercrime, the experts rate the country's scores in three dimensions (influence, professionalism, and technicality), and then calculate the average score of each country for each crime type. These scores will be weighted and adjusted according to the number of nominations to ensure that countries with fewer nominations do not rank high due to high scores. Next, the average scores of each country across the five types of cybercrime are calculated to obtain the country's total cybercrime type score. Finally, the scores are further adjusted based on the total number of nominations from all countries to ensure that the final score range is between 0 and 100. The formula

for the WCI score is as follows:

$$\mathrm{WCI}_{\mathrm{overall}} = \left(\frac{1}{5}\right)\sum_{i=1}^{5}\left[\left(\frac{1}{\mathrm{nominations}}\right)\sum_{j=1}^{\mathrm{pominations}}\frac{(I_j+P_j+TS_j)}{3}\times\left(\frac{\mathrm{nominations}_i}{92}\right)\times 10\right]\times\left(\frac{\mathrm{nominations}_i}{460}\right)\times 10 \quad (1)$$

(1) Country: assessed in terms of its influence (I), professionalism (P), and technological sophistication (TS).

(2) Nominations: the number of nominations for a certain country in all categories.

(3) Maximum nominations per type: 4.92, representing the maximum number of nominations for each type of cybercrime.

(4) Maximum nominations overall: 460, representing the maximum number of nominations for all types of cybercrimes (92 nominations for each country for each type of cybercrime, totaling 5 types of cybercrimes).

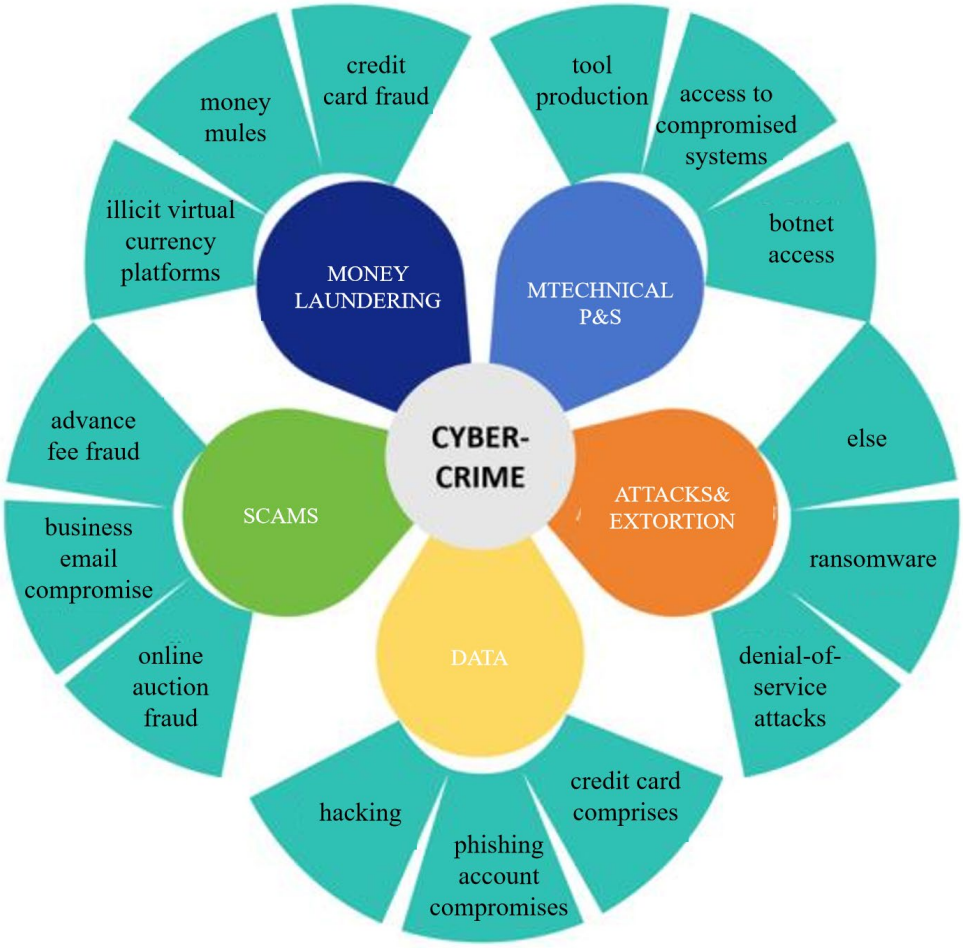

Figure 1 Overview of Global Cybercrime Index Categories.

The final WCI score represents the intensity and danger of cybercrime activities in various countries, providing a reference for formulating more precise cybersecurity prevention measures.

3.1.2. Global Cybercrime Risk/Exposure Index

This study refers to existing practices and first collects data on five types of attack rates, namely Malware Encounter Rate, Ransomware Encounter Rate, Cryptocurrency Mining Encounter Rate, Drive-by Download Page Encounter Rate, and Cloud Provider Related Incoming Attacks, from the latest Microsoft Digital Defense Report, and sorts them from low to high according to the degree of exposure. Secondly, it reverses the ranking of the Global Cybersecurity Index obtained from ITU. Then, each country's respective ranking is summed and divided by 290 to calculate the exposure scale from 0 to 1 (low to high).

The formula for obtaining the global cybercrime exposure score is:

$$CEI = \frac{\sum_{i=1}^{5} R_i + R_{commitment}}{290} \quad (2)$$

R: The rankings of malicious software, ransomware, cryptocurrency mining, drive-down pages, and cloud provider-related inbound attacks are sorted from high to low based on the degree of exposure, with larger numbers indicating higher exposure. $R_{commitment}$ represents the ranking of the degree of cybersecurity commitment, sorted from high to low based on the degree of commitment. Countries with higher commitment rank lower, indicating lower exposure. 290 is the maximum total sum for all countries' rankings (Afghanistan has the highest total sum).

The global cybercrime exposure score is calculated using the following formula, which reflects the degree of global cybercrime risk. The higher the score, the higher the degree of potential crime in the region, and corresponding measures for cybercrime prevention and blocking should be taken.

3.2. Distribution Characteristics and Analysis of Cybercrimes

Clustering Model: The first category is the high-incidence distribution area; the second is the national characteristics of successful organized crime. Through the analysis of these different national characteristics, factors that may influence the rate of cybercrime are identified.

K-means Clustering Algorithm: Collect data related to cybercrime from various countries, including the number of crimes, the amount of losses, successful criminal cases, etc. Conduct K-means clustering based on these data to obtain different regions with a high incidence of cybercrime. Analyze the national characteristics of these high-incidence regions and identify the common features of these regions, such as economic development level, internet penetration rate, legal enforcement capacity, etc. Identify the key factors influencing the distribution of cybercrime and the effectiveness of prevention, such as technical, legal, and socio-economic factors.

Based on these data, K-means clustering is conducted to obtain different crime-prone areas. The national characteristics of these high-crime areas are analyzed to identify the common features of these areas, such as economic development level, internet penetration rate, and legal enforcement capacity.

According to the WCI (World Cybercrime Index) heat map, we can now analyze the impact of policies, economy, education, and communication technologies on cybercrime. Countries or regions that have formulated stricter policies (such as those marked on the map that implement the "Global Personal Information Protection Regulations") tend to have lower rates of cybercrime. This supports the argument that government policies play a key role in reducing cybercrime. For instance, Western European countries, especially those in the European Union, may lower the rate of cybercrime due to strict policies and regulations. This might be the reason why regions such as Germany and Scandinavia show lower rates of cybercrime on the heat map, which is related to the implementation of policies like GDPR, which can ensure better management of data security and privacy.

The economic conditions of the regions also have a significant impact. As shown in the Figure 2, the crime indices in prosperous regions such as North America, parts of Western Europe, and East Asia tend to be relatively high. This might be related to the prevalence of high-value digital assets and network data accumulated by enterprises, as these assets and data are more likely to be exploited by cybercriminals.

Figure 2 Global Distribution of Cybercrime Index (WCIWCI_Overall) by Country.

In terms of education, regions with higher educational attainment, especially those where the rate of higher education enrollment is high and the proportion of the labor force with higher education is large, tend to have a larger population with greater awareness and stronger moral consciousness. This can reduce the likelihood of engaging in cybercrime, as individuals are more likely to be aware of the legal consequences.

Finally, communication and technological infrastructure are indispensable facilitators for cybercrime. Regions with more advanced communication networks and digital technologies may become hotspots for cybercrime activities. This is especially true for areas where network devices and services are widely available, providing more opportunities for cybercriminals to exploit vulnerabilities.

3.3. Relationship between Reporting Rate and Success Rate of Criminal Acts

3.3.1. Cybercrime Metrics: Reporting, Prevention, and Success Rates

This study utilizes complaint cases reported by the Internet Crime Complaint Center (IC3) and combines data from online forums and web pages to complete the global network crime complaint volume for 2023. After logarithmic processing and percentage calculation, the complaint rates of network crimes in various countries are obtained. Moreover, given the increasingly serious cybersecurity threats, the KYC (Know Your Customer) process can enhance security, ensure compliance, prevent abuse, and contribute to improving the overall user experience. Taking proactive measures such as locking service accounts is crucial for preventing malicious activities. Due to the lag in cybersecurity investment—for example, only 50% of American enterprises have complete cyber insurance, and there is a shortage of cybersecurity professionals globally—enterprises must rely on effective security protocols. Therefore, based on the KYC process, to construct a calculation formula for network risk prevention, we can comprehensively evaluate the risk of accounts by considering multiple security factors. Assuming that the security of an account is affected by multiple factors, including KYC compliance, trial abuse, suspicious transactions, device verification, and the completeness of security protocols. Through the weighted combination of these factors, we can calculate a risk score, RiskScore, and based on this score, determine whether to take preventive measures such as locking the account. When the calculated RiskScore exceeds a preset threshold T, the account will be locked to prevent potential risks and abuse. The formula for network risk

prevention is obtained as follows:

$$RiskScore = \sigma(\alpha_1 \cdot X_1 + \alpha_2 \cdot X_2 + \alpha_3 \cdot X_3 + \alpha_4 \cdot X_4 + \alpha_5 \cdot X_5) \quad (3)$$

$\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ represent the weights of each factor, indicating the degree of influence of each factor on the overall risk score. X_1, X_2, X_3, X_4, X_5) respectively represent the status of KYC compliance, trial abuse, suspicious transactions, device verification, and security protocol, with values of 0 or 1. A value of 0 indicates no risk, and 1 indicates the existence of risk.

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (4)$$

This is the Sigmoid function. It maps the weighted values to a range of 0 to 1, which represents the probability that an account will be locked.

Finally, due to the lack of complete statistics on subjectivity, this study considers using the global malicious network attack score as an instrumental variable for the success rate of cybercrime. The higher the malicious network attack score, the higher the success rate it indicates. By using disclosed data from the Internet Crime Complaint Center, after matching domain names with countries and regions through the query matching of WHOIS information method from the TLD Phishing Table, Registrar Phishing Table, and Hosting Networks Phishing Table, we obtained more complete data. We cleaned the data and merged it according to the corresponding fields, aggregating the relevant data of each TLD, and calculated its phishing attack score. We calculated the average malicious phishing attack score, average malicious software activity score, and average malicious spam activity score for each country. Finally, we obtained the malicious network attack score through principal component analysis.

Select the eigenvector corresponding to the largest eigenvalue as the new data projection direction. Finally, the selected eigenvector will project the standardized data to the new space to obtain the representation of malicious network attacks in the principal component space.

3.3.2. Hypothesis Testing

This study employs five ordinary least squares (OLS) regression models to validate the following five questions: (1) Does network crime reporting reduce the success rate of crimes? (2) Can network crime reporting prevent network crimes? (3) Can network crime reporting reduce the success rate of network crimes by increasing the risk associated with such crimes? (4) What is the impact of the relationship between a country's network crime environment and network crime reporting on the success rate of network crimes? The model formulas and test results are presented in Table 1 and Table 2. The analysis was conducted using a sample size of 27,315 observations. The robustness of the results was tested through 500 simulations.

Table 1 Ols regression analysis of network crime reporting and success rate.

Variables	(I) success rate	(II) complaint volume	(III) success rate	(IV) success rate	(V) success rate
complaint	-0.638***		-0.5991***	-0.593***	-0.378***
volume	(11.841)		(-3.463)	(-7. 137)	(-5.310)
inhibition rate		0.3644***	-0.2791***		
		(3.543)	(-3.463)		
interaction				0. 1433**	
term1				(8.674)	
interaction					0.464***
Controls	control	control	control	control	control
Adjusted R-squared	0.4009	0. 1222	.874	0.127	0.4631
n	167	167	167	167	167

Table 2 Mediation analysis results.

	Estimate	95% CI Lower	95% CI Upper	p-value	
ACME	0.007	0.004	0.010	<2e-16	***
ADE	0.006	0.002	0.010	0.02	**
Total Effect	0.013	0.002	0.020	<2e-16	***
Prop. Mediated	0.523	0.363	0.870	<2e-16	***

The results indicate that when the network crime index (Complaint Volume) is elevated, network crime reporting has a negative effect on the success rate of crimes (coefficient = ; p-value =), suggesting that an increase in reporting contributes to a decrease in the success rate of network crimes. Regression analysis further reveals that the rise in reports has enhanced the block rate (Block Rate), demonstrating that reporting can effectively prevent network crimes. Additionally, the third figure illustrates the role of network crime reporting in reducing the success rate of crimes by increasing the block rate, indicating that reports not only directly prevent crimes but also further decrease the success rate of network crimes by enhancing the block rate. Thus, in this context, network crime reporting can not only effectively prevent network crimes but also reduce their success rate by improving risk prevention capabilities.

Furthermore, to explore the network crime environments in which reporting is most beneficial for crime prevention, a consistency analysis was conducted. The three-dimensional response surface plot provides insights into these conditions.

The three-dimensional response surface and cross-sectional diagrams reveal that the effectiveness of reporting is more pronounced when network criminal activities are less frequent. This suggests that reporting can facilitate the rapid detection and prevention of criminal acts in such environments. Conversely, in high-risk network criminal environments, criminal activities tend to be more covert and extensive. While reporting does have some impact, it may be mitigated by various complex factors, such as countermeasures and anonymity, which can undermine its effectiveness in reducing network crimes in areas characterized by severe and high-risk conditions. Therefore, relying solely on non-governmental spontaneous reporting may not be sufficient to combat network crimes in these serious and high-risk areas. It is necessary to integrate additional technical measures for comprehensive protection.

4. Effectiveness Assessment of Cybersecurity Policies

This study aims to explore the relationship between cybersecurity policies and cybercrime, evaluate the effectiveness of these policies across various countries, and identify the key components and an ideal or generalized framework for cybersecurity policies.

4.1. Policy Effectiveness Analysis: Constructing the National Cybersecurity Index

Drawing on existing research and available data, this study constructs the National Cybersecurity Index (NCI) as a pivotal metric for assessing the efficacy of cybersecurity policies. Leveraging the previously calculated Global Cybercrime Index and Risk Index, the Analytic Hierarchy Process (AHP) is applied to ascertain the relative weights of each contributing factor. This process facilitates the computation of NCI values for a range of countries. The NCI encompasses dimensions including the rigor of policy enforcement, the impact of economic investments, the influence of education and workforce quality, and the level of communication and technical support, among others. A higher NCI value signifies more effective cybersecurity policies and a more robust cybersecurity posture for a given country.

4.1.1. Construction of Hierarchical Structure Model

Utilizing the National Cybersecurity Index (NCI) as the objective layer, the hierarchical structure model is formulated with criterion layers comprising policy enforcement intensity, economic investment impact, the influence of education and labor force quality, and the level of communication and technical support. Beneath each criterion layer lie the indicator layers, which consist of specific

sub-factors that serve to measure the respective criteria. This structured approach delineates a comprehensive framework for assessing a nation's cybersecurity profile by dissecting the multifaceted concept of cybersecurity into quantifiable and assessable elements.

4.1.2. Construction of Judgment Matrix

Employing the expert scoring method, judgment matrices are developed for the criterion layer in relation to the target layer and for the indicator layer in relation to the criterion layer. Experts provide pairwise scores for the four factors within the criterion layer—policy enforcement intensity, economic investment effect, influence of education and labor force quality, and communication and technical support level—based on their relative importance to the National Cybersecurity Index (NCI). The 1-9 scale method is used for scoring, where 1 indicates equal importance, 3 indicates slight importance, 5 indicates strong importance, 7 indicates very strong importance, and 9 indicates extreme importance, with 2, 4, 6, and 8 as intermediate values.

Following this methodology, the judgment matrix for the criterion layer is then constructed as follows:

$$A = \begin{pmatrix} 1 & a_{12} & a_{13} & a_{14} \\ a_{21} & 1 & a_{23} & a_{24} \\ a_{31} & a_{32} & 1 & a_{34} \\ a_{41} & a_{42} & a_{43} & 1 \end{pmatrix} \tag{5}$$

4.1.3. Hierarchical Single Sorting and Consistency Check

The first step involves calculating the weight vector by determining the maximum eigenvalue of the judgment matrix and its corresponding eigenvector. This process involves normalizing the eigenvector to obtain the relative weights of each factor. Specifically, for the judgment matrix A, the maximum eigenvalue and the eigenvector W, which consists of components w1, w2, w3, and w4, are computed. Subsequently, the eigenvector is normalized to yield the weight vector w, composed of the same components w1, w2, w3, and w4.

The second step is a consistency check to ensure the reliability of the judgment matrix. This is achieved by calculating the consistency index CI, which is dependent on the order of the judgment matrix. Additionally, the random consistency index RI is used, which varies according to the order of the judgment matrix and is typically obtained from a table. The consistency ratio is then calculated. If the consistency ratio CR is less than 0.1, the judgment matrix is deemed to have satisfactory consistency. If the ratio exceeds 0.1, adjustments to the judgment matrix are necessary to improve its consistency.

4.1.4. Hierarchical Overall Ranking and Consistency Check

To calculate the overall ranking weights of the hierarchy, we begin with the weights of each factor obtained from the single-level ranking. Using these weights, we then determine the overall ranking weights of each factor in the indicator layer with respect to the target layer. This involves multiplying the weight vector of the criterion layer for the target layer by the weight vectors of each criterion layer. The resulting overall ranking weight vector for each factor in the indicator layer is derived from this multiplication process.

For the second step, an overall sorting consistency test is performed. This test follows a method similar to that used for the consistency test of single-level sorting. It involves calculating the consistency index and the consistency ratio for the overall sorting of the hierarchy. Based on these calculations, a consistency test is conducted to ensure that the judgments made at each level of the hierarchy are consistent with each other. This step is crucial for validating the reliability of the overall ranking and determining if any adjustments are needed to improve the consistency of the hierarchical structure.

4.1.5. Overall Sorting Consistency Test

The first step involves collecting data for each country across all levels of the factors and

conducting data standardization processing. Given the variance in scales and value ranges among different indicators, it is essential to standardize the data to ensure comparability. In this study, the range-based standardization method is employed to normalize the data across different indicators.

Following the standardization, the National Cybersecurity Index (NCI) values for each country are calculated. This calculation is based on the weights assigned to each factor from the overall ranking of the hierarchical model and the standardized data. The NCI for each country is determined by aggregating the weighted standardized data of all indicators, with each weight corresponding to the total ranking weight of an indicator factor in the indicator layer relative to the target layer. This comprehensive calculation provides a measure of the cybersecurity index for each country, reflecting the collective impact of various factors on the country's cybersecurity posture.

4.2. General Framework for Identifying the Universality of Cybersecurity Policies: Supply Chain Game Model

4.2.1. Model Setup

(1) Establishing Game Entities and Their Interactions

A game-theoretic model is developed to encapsulate the dynamics between the state (policymakers) and cybercrime groups. The state's role involves the creation and enforcement of cybersecurity policies, which includes determining the rigor of these policies, the extent of resource allocation, and the level of information transparency. In response, cybercrime groups adjust the scale and frequency of their illicit activities based on the state's policy environment, the costs associated with their crimes, and the anticipated returns.

(2) Incorporating Additional Considerations

Resource Dependency of Cybercrime Groups: The reliance of cybercrime groups on specific network resources, such as sophisticated technologies, financial capital, and skilled personnel, influences their decision-making processes. High dependency on these resources makes cybercrime groups particularly vulnerable to the state's regulatory policies targeting such assets, thereby exerting a substantial influence on their operations.

Policy Implementation and Power Structure: The capability of the state to effectively implement policies and allocate resources in the cybersecurity domain is crucial. It establishes the state's strategic advantage in its interaction with cybercrime groups. Greater policy implementation power can effectively induce behavioral adjustments within cybercrime groups.

Environmental Contingency: External factors, including the global cybersecurity landscape, the pace of domestic network technology advancements, and societal pressure through public opinion, can alter the behavioral patterns of both the state and cybercrime groups. For instance, during periods of heightened global cybercrime activity, the state may intensify its policy initiatives, prompting cybercrime groups to exercise greater caution in their activities.

(3) Analysis of Information Structure

The quality of national cybersecurity information disclosure is quantified by a parameter μ, where μ is less than or equal to 1. Perfect transparency is indicated when μ equals 1; partial transparency or the presence of concealment is suggested when $0 < \mu < 1$; reliance on prior experience occurs when μ is 0; and information catering behavior is inferred when μ is less than 0. Cybercrime groups assess the risks associated with criminal activities based on the information disclosed by the nation and their own intelligence gathering.

The Resource Dependence Coefficient is determined by calculating the proportion of resources Rc that a cybercrime group can access relative to the resources Rn invested by the country in cybersecurity, expressed as Rc/Rn. A lower coefficient signifies a greater dependency of the cybercrime group on the nation's resources, leading them to closely monitor shifts in national policies.

(4) Explanation of Basic Assumptions

The "Demand" Function of Cybercrime Behavior: Let D represent the total potential number of cybercrime incidents, Q denote the amount of national investment in cybersecurity resources, and θ (defined as $\theta = em/em0$) indicate the policy strictness, where em is the reduction in criminal behavior following the implementation of policies, and em0 is the anticipated volume of crime in the absence

of policy intervention. The "demand" function for cybercrime behavior is constructed based on these variables.

$$D = aQ(1-\gamma) - \beta_1 p_1 + \beta_2 p_2 + n_0\theta \quad (6)$$

In this context, parameters α, γ, β_1, and β_2 represent the revenue generated from network crimes, reflecting the influence of various external factors on the crime proceeds. When considering the cost and benefit factors, it is important to note that the network crime group incurs a cost C_p when engaging in criminal activities. This cost encompasses technical investments, risk-related expenses, and other factors. Additionally, the state's policy formulation and implementation also involve costs C_n, which may include expenditures on human resources, technological research and development, and equipment purchases.

The cost associated with the implementation of state policies is related to the stringency of the policies and can be expressed as $k_1\theta^2$, where θ represents the policy strictness. The cost for the network crime group to adapt to policy changes is given by $k_2(\theta - \theta_0)^2$, with θ_0 denoting the minimum policy strictness required by the state. Furthermore, the cost of state resource investment is expected to decrease with advancements in technology and international cooperation, which can be represented by the term Dt.

To account for the influence of multiple variables, the function is adjusted to incorporate policy, economic, education and labor force, and communication and technology variables. This adjustment allows for a more comprehensive analysis of the factors affecting network crime revenue and the costs associated with crime prevention and policy implementation.

5. Optimization of Objectives Setting

The national goal is to maximize the National Cybersecurity Index (NCI), which is aimed at reducing both the potential number of cybercrime incidents and the costs associated with national cybersecurity. The costs incurred by the nation, denoted as Cn, include multiple components: the cost of information disclosure (Cμ), which involves expenses related to transparency and the dissemination of cybersecurity information; the cost of policy implementation (Cθ), which covers the expenses for enacting and enforcing cybersecurity policies; the cost of resource acquisition (Cr), which pertains to the resources needed to combat cybercrime; the cost of motivating criminal groups to alter their behavior (Cw), which includes efforts to deter and dissuade cybercriminal activities; and the cost of responding to changes in the external environment (Ce), which accounts for the nation's adaptability to evolving cybersecurity threats.

Conversely, the objectives of network criminal groups are focused on maximizing their illicit gains from cybercrime activities. Their costs, represented as Cc, include a variety of expenses such as the costs associated with acquiring information, investing in criminal techniques, and the costs of evading policy measures. These costs reflect the resources and efforts that cybercriminals must allocate to carry out their activities successfully.

5.1. Event Process and Equilibrium Analysis

(1) Event Process Description

In a given period, the country develops cybersecurity policies based on the current cybersecurity landscape and its objectives. It decides on the level of resource investment, policy strictness, and information disclosure. Upon receiving this information, cybercrime groups evaluate the costs and benefits of their actions and choose the scale and frequency of their criminal activities. Subsequently, the country and the cybercrime groups adjust their strategies based on the outcomes of their interactions and monitoring data.

(2) Equilibrium Analysis

The country aims to maximize its cybersecurity index, considering various factors such as resource allocation and policy strictness. Cybercrime groups, on the other hand, aim to maximize their gains while considering the country's policies and their operational costs. By analyzing the decisions of both parties, we can determine the optimal strategies for each, using methods like reverse induction

to find the equilibrium points.

(3) Solution and Analysis

The goal is to identify the optimal policy combinations that maximize the National Cybersecurity Index (NCI). This involves simulating different scenarios to understand how various policy combinations affect the NCI. The analysis considers factors like legal improvements, economic investments, education levels, and technological advancements to determine the most effective strategies in different cybersecurity environments.

6. Conclusion

This study has provided a comprehensive analysis of the global distribution of cybercrime and the effectiveness of cybersecurity policies through the application of the K-means clustering method and the Analytic Hierarchy Process (AHP) model. The K-means clustering effectively revealed distinct patterns in cybercrime distribution, highlighting regions with high incidence rates and identifying key socio-economic and technological factors that contribute to vulnerability. Meanwhile, the AHP model successfully constructed the National Cybersecurity Index (NCI), offering a robust metric to evaluate policy effectiveness across different countries. This dual-method approach not only enhances our understanding of the complex interplay between cybercrime and cybersecurity policies but also provides valuable insights for policymakers aiming to optimize their strategies.

However, there are areas for improvement. The current models, while comprehensive, could benefit from the integration of more granular and real-time data to enhance accuracy and responsiveness. Additionally, future research could explore the incorporation of emerging technologies, such as artificial intelligence and machine learning, to further refine the predictive capabilities of the models. Moreover, expanding the scope to include a broader range of socio-economic indicators and international cooperation metrics could offer a more nuanced understanding of the global cybersecurity landscape. By addressing these areas, future studies can build upon the foundation laid by this research to develop even more effective tools for combating cybercrime and enhancing global cybersecurity.

References

[1] Sabillon R, Cano J J, Serra-Ruiz J. Cybercrime and cybercriminals: A comprehensive study[J]. International Journal of Computer Networks and Communications Security, 2016, 4 (6), 2016.

[2] Grispos G. Criminals: Cybercriminals[M]//Encyclopedia of Security and Emergency Management. Cham: Springer International Publishing, 2021: 84-89.

[3] DeTardo-Bora K A, Bora D J. Cybercrimes: An overview of contemporary challenges and impending threats[J]. Digital Forensics, 2016: 1.

[4] Tsai C H, Zdravkovic J, Stirna J. Modeling digital business ecosystems: a systematic literature review[J]. Complex Systems Informatics and Modeling Quarterly, 2022, (30): 1-30.

[5] Sviatun O, Goncharuk O, Roman C, et al. Combating cybercrime: economic and legal aspects[J]. WSEAS Transactions on Business and Economics, 2021, 18: 751-762.

[6] Mijwil M M, Aljanabi M, ChatGPT C. Towards artificial intelligence-based cybersecurity: The practices and ChatGPT generated ways to combat cybercrime[J]. Iraqi Journal For Computer Science and Mathematics, 2023, 4(1): 8.

[7] Shafqat N, Masood A. Comparative analysis of various national cyber security strategies[J]. International Journal of Computer Science and Information Security, 2016, 14(1): 129-136.

[8] Mishra A, Alzoubi Y I, Anwar M J, et al. Attributes impacting cybersecurity policy development: An evidence from seven nations[J]. Computers & Security, 2022, 120: 102820.

[9] Darko A, Chan A P C, Ameyaw E E, et al. Review of application of analytic hierarchy process

(AHP) in construction[J]. International journal of construction management, 2019, 19(5): 436-452.

[10] Vaidya O S, Kumar S. Analytic hierarchy process: An overview of applications[J]. European Journal of operational research, 2006, 169(1): 1-29.

[11] Mishra A, Alzoubi Y I, Gill A Q, et al. Cybersecurity enterprises policies: A comparative study[J]. Sensors, 2022, 22(2): 538.

[12] Bartholomae F. Cybercrime and cloud computing. A game theoretic network model[J]. Managerial and Decision Economics, 2018, 39(3): 297-305.

[13] Merrick K, Hardhienata M, Shafi K, et al. A survey of game theoretic approaches to modelling decision-making in information warfare scenarios[J]. Future Internet, 2016, 8(3): 34.