mathematics

MDPI

Article

Information Leakage Detection and Risk Assessment of Intelligent Mobile Devices

Xiaolei Yang, Yongshan Liu * and Jiabin Xie

School of Information Science and Engineering, Yanshan University, Qinhuangdao 066000, China; yangxl@stumail.ysu.edu.cn (X.Y.); ean@stumail.ysu.edu.cn (J.X.)
* Correspondence: jsjbs0019@163.com

Abstract: (1) Background: Smart mobile devices provide conveniences to people's life, work, and entertainment all the time. The basis of these conveniences is the data exchange across the entire cyberspace, and privacy data leakage has become the focus of attention. (2) Methods: First, we used the method of directed information flow to conduct an API test for all applications in the application market, then obtained the application data transmission. Second, by using tablet computers, smart phones, and bracelets as the research objects, and taking the scores of senior users on the selected indicators as the original data, we used the fusion information entropy and Markov chain algorithm skillfully to build a data leakage risk assessment mode to obtain the steady-state probability values of different risk categories of each device, and then obtained the entropy values of three devices. (3) Results: Tablet computers have the largest entropy in the risk of data leakage, followed by bracelets and mobile phones. (4) Conclusions: This paper compares the risk situation of each risk category of each device, and puts forward simple avoidance opinions, which might lay a theoretical foundation for subsequent research on privacy protection strategies, image steganography, and device security improvements.

Keywords: directed information flow; information disclosure; information entropy; Markov; risk assessment

MSC: 60J20; 94A17

Citation: Yang, X.; Liu, Y.; Xie, J. Information Leakage Detection and Risk Assessment of Intelligent Mobile Devices. *Mathematics* **2022**, *10*, 2011. https://doi.org/10.3390/math10122011

Academic Editor: Daniel-Ioan Curiac

Received: 6 May 2022
Accepted: 9 June 2022
Published: 10 June 2022

1. Introduction

With the rapid development of science and technology, the electronic platform is becoming more and more intelligent and mobile, which has brought great convenience to people's life. Today, with the prevalence of big data, the data itself are also spreading along the trend of large depth, high production speed, wide dimensions, and low density. At the same time, the means for hackers to steal information is also powerful, resulting in the outflow of a large number of personal privacy data [1]. Information leakage has become a hot topic in today's cyberspace. How to detect, describe, and even protect privacy has become the focus of the netizens' close attention.

In 2018, the personal information of 87 million Facebook users was leaked. In September of the same year, the information of another 30 million users was leaked due to hacker attacks, and the data of 68 million users were leaked due to software vulnerabilities on 14 December. On 10 January 2019, Bob Diachenko, a hackenproof security researcher, found that the detailed resume information of more than 202 million Chinese job seekers in the mongodb database was published online, which was suspected to be leaked by third-party applications. It is reported that the 202 million resumes stored in this database contain 202,730,434 records with very detailed information including the applicant's name, height, weight, address, date of birth, telephone number, email address, political orientation, skills, work experience, salary expectation, marital status, driver's license number, professional

experience, and career expectation, totaling 854 gb. In August 2020, a logistics company in Hebei Province, China reported that its employee account was monitored by the company's logistics risk control system for the illegal inquiry of the waybill number information of non-local outlets, resulting in the possible disclosure of a large number of the customers' privacy information. On the evening of 15 March this year, the annual "15 March" party was broadcast on the central finance and economics channel. The link of "improving digital rules and building Internet economic confidence" exposed the problem of personal privacy leakage in enterprises: Zhilian recruitment failed to pass the examination of enterprises, resulting in a large number of downloads of the resumes of job seekers. As a result, there are many risks of private information leakage around us.

"Privacy computing theory" first appeared in 1999. It pointed out that information will be leaked only when device users think that the benefits are equal to the risks [1,2]. Guo Yu's research showed that data information disclosure positively affected the privacy information disclosure behavior, perceived mobile learning profitability, and privacy control while self-efficacy positively affected the privacy information disclosure intention, and the perceived mobile learning risk negatively affected the users' privacy information disclosure intention [3]. By studying the privacy information disclosure behavior and protection of mobile device users, Xiong Jian showed that the factors of the perceived benefits and perceived risks had a strong impact on the users' self-perceived willingness [4]. Wang Kan used comprehensive fuzzy evaluation to evaluate the risk of data leakage in a transaction, in which the risk factors included network access control, network application protocols, firewalls, and identity authentication [5]. Zhao Zhuohe found that the wireless network used by mobile devices was easy to intercept, resulting in important information and data being stolen [6]. Li Yanhui believed that the wireless network is open and easy to obtain its internal structure, so as to obtain important data nodes for targeted interception [7]. Xu Jiale suggested that the social network or platform failed to strictly control the enterprise qualification, resulting in the platform's inability to trace the source of information leakage [8]. Makhdoom believed that anonymous encryption could make greater efforts to ensure that receipts were not disclosed [9]. To sum up, for smart mobile devices, the risk of user information disclosure is distributed in all corners of cyberspace. Although there are many studies on the risk of privacy disclosure, only a few can comprehensively and in detail describe the risk factors of privacy disclosure and evaluate the risk of the information disclosure of tablets, smartphones, and bracelets. Therefore, this paper subdivides and expands the risk factor indicators considered in the above articles, and finally combined them into five categories and 24 risk indicators to comprehensively evaluate the risk of the privacy disclosure of tablets, smartphones, and bracelets.

First, based on the directed information flow detection risk application, this paper constructed an information flow model to track and analyze the privacy points in real time. Then, it summarizes the various risk factors of intelligent mobile devices in wireless networks, selects the risk indicators, and constructs an evaluation model based on information entropy and Markov chain. Finally, according to the evaluation results, targeted preventive measures will be issued and implemented.

2. Malicious Application Detection Based on Directed Information Flow

2.1. Basic Theory

Information flow is a classic method to detect the information leakage of risky applications. This method was born in 1976 and is based on Denning's grammatical information flow analysis:

$$FM = \langle N, P, SC, \oplus, \rightarrow \rangle \quad (1)$$

where N is the set of some logical elements (code segments, variables, etc.) in the system; P is the collection of processes and the response subject of information flow; SC is the collection of safety levels, which is used to judge whether the operation behavior is legal; $\oplus$ is the operational supremum of the security level, and the result is the minimum common

upper bound of security levels A and B. This indicates the flow direction of the information flow, which means that the information in A is allowed to flow to B [10,11].

The syntax information flow detection steps are shown in Figure 1.

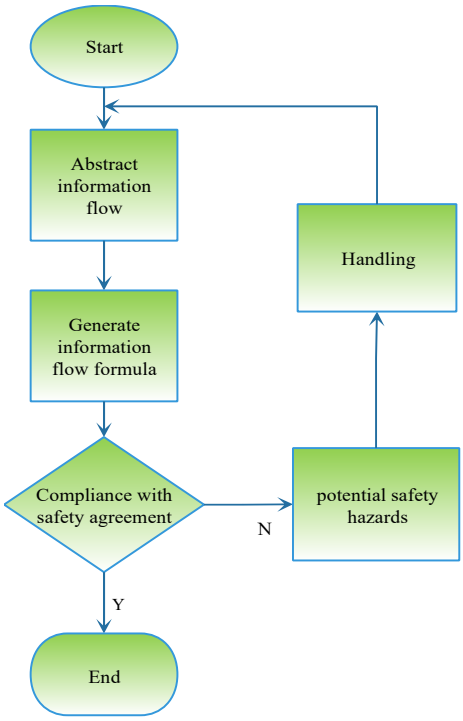

Figure 1. The flow chart of the information flow detection.

In addition to malicious applications, privacy information leakage may also occur in various stages of big data computing. As shown in Figure 2, under the cloud platform-based big data computing, privacy leakage may occur during the data transmission from the application to the cloud service provider, the cloud platform computing process, and the cloud platform data output phase. Therefore, we focused on detecting private data, and whether this is directly transmitted to the external cyberspace, and if so, if the application software is regarded as software with the risk of privacy leakage.

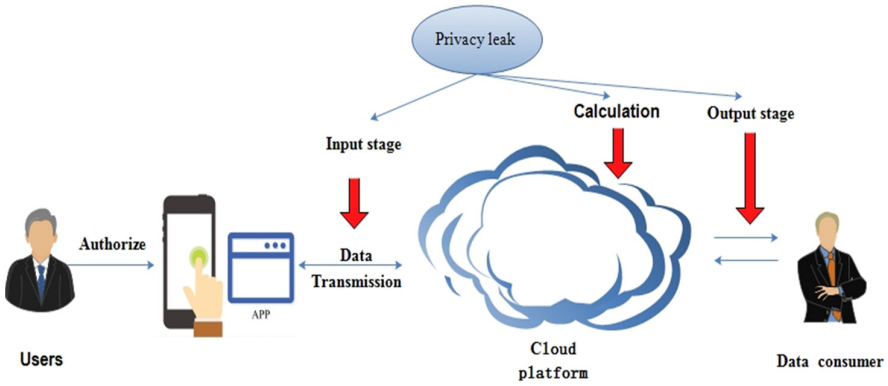

Figure 2. The cloud platform-based big data computing environment.

The method can roughly be divided into three steps: first, abstract the information flow, analyze the object source code, and extract the idiom meaning of the information manifold in each line of code [2,12]. The second is to form the information flow formula, which requires the design of an information flow strategy [13]. Finally, the formula is used to verify whether the information flow formula complies with the security level agreement.

If it does, it indicates that the formula is correct, otherwise, it indicates that there is a potential security hazard. In order to avoid the problem of false alarms, the verification is carried out again according to the information flow method after appropriate treatment. If it fails to pass the security level agreement many times, it will be recognized as a potential safety hazard.

Directed information flow: According to the privacy point dataset, analyze all function calls in the Java source code and read/write privacy data, and finally form an information flow model. If private information eventually flows to the outside cyberspace, it is considered that there is a privacy disclosure [10,14]. For example, if the top function is a network connection function and passes private data as connection parameters, or the top function is a SMS sending function and sends private data as SMS content, it is considered that the application is a malicious application, which will lead to the theft of the users' private information [15].

The output module arranges, counts, analyzes, and outputs the detection results and finally forms a complete analysis report to list the specific contents of risky applications.

$$AM = \langle \mathrm{L}, O, F, \rightarrow \rangle \tag{2}$$

where AM is the information flow model; L is the set of leakage points; O is the set of all external interaction functions in all devices; and F is the set of calling and operating functions.

$$f \rightarrow \mathrm{l}, f \in F, \mathrm{l} \in \mathrm{L} \tag{3}$$

If any privacy point accesses the function call, it forms a directed information flow:

$$f_n \rightarrow \cdots \rightarrow \cdots f_2 \rightarrow f_1 \rightarrow \mathrm{l}, f_i \in F, \mathrm{l} \in \mathrm{L} \tag{4}$$

Moreover, $f_n \in O$ indicates that the privacy has been compromised, and the application is identified as a suspected malicious application output.

For multiple branch information flows:

$$\begin{array}{c} f_n^{'} \rightarrow \cdots \rightarrow \cdots f_2^{'} \rightarrow f_1^{'} \rightarrow \mathrm{l}, f_i^{'} \in F, \mathrm{l} \in \mathrm{L} \\ \cdots \end{array} \tag{5}$$

As long as one item is satisfied, $f_n^x \in O$, it is also recognized as a malicious application.

2.2. Network Environment

The network environment of an application or app is divided into two parts, one is the data flow between the hardware framework and the external network, and the other is the data flow from the operating system and software itself to the external network.

The network environment detection of intelligent mobile devices is carried out in the process of data exchange between the software and hardware of the device and the external network (see Figure 3 for the specific detection framework). Among them, the hardware detection mainly involves four parts: Event Signature, Event Classification, Event Input, and Event Detection [16]. Event Signature is an important part of the detection of an information leakage event and is trained according to the historical data. The target value is whether it is defined as a privacy event. After the training, it uses machine learning to classify the unknown data to be detected. Event Input is the newly generated data sample to be tested. Software testing mainly involves API (Application Programming Interface) Acquisition, APP Reverse, and API Testing [17]. The principle of software detection is to obtain the API containing privacy features from the data flowing out of the device, find the parameters or methods to generate privacy data according to the reverse tools, and detect whether there is any leakage of the tag information by changing or marking the parameter information in the software.

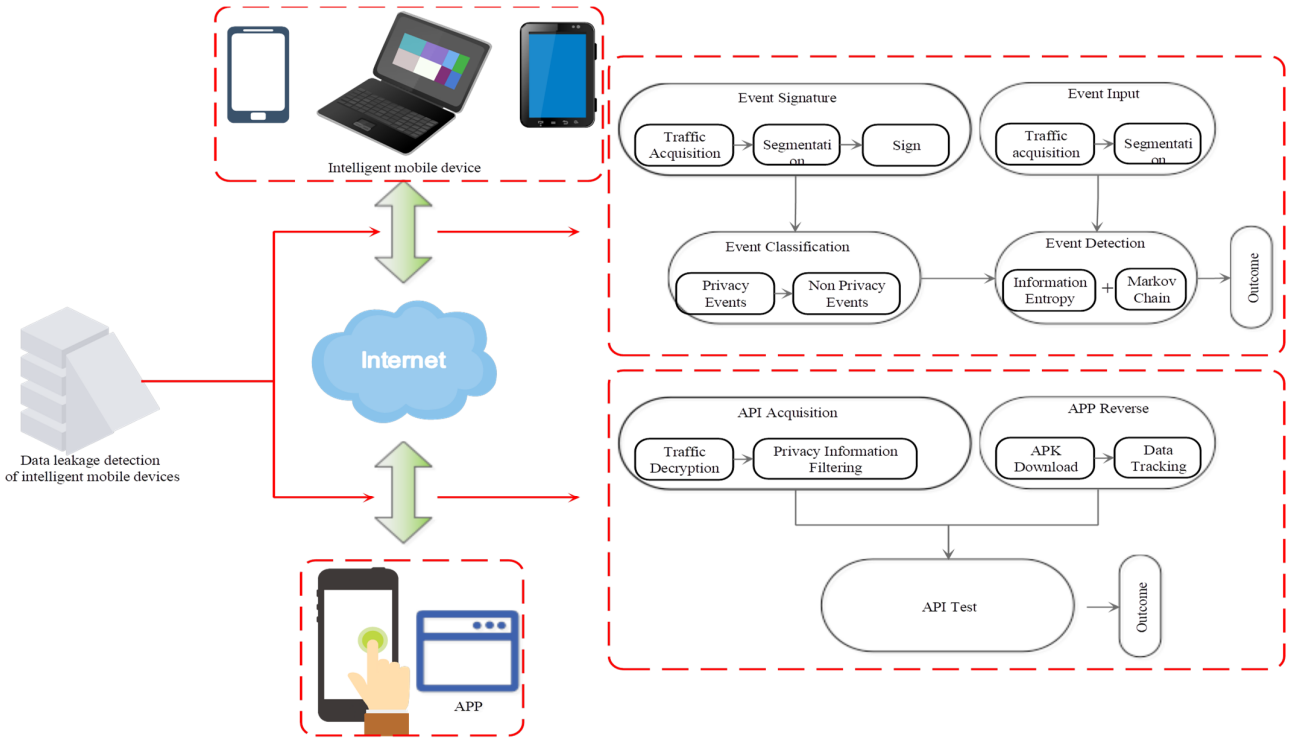

Figure 3. The data leakage detection process of intelligent mobile devices.

In the process of the detection of information leakage from mobile devices to the external cyberspace, the characteristics of risky applications and high-risk API source codes are often used (see Tables 1 and 2 for details).

Table 1. The risky application characterization table.

Application Program	Risky Application Characterization
Message	Obtain the content of message, sending and receiving time and SMS records
Contacts	Obtain address book information
Instant Messaging	Obtain communication software information, such as WeChat record
Browser	Obtain browser access history, tag data, etc.
Call Log	Obtain call record, call time, call frequency
Social Networks	Obtain social app data, such as takeout data and likes
Position	Obtain position information, motion trajectory

Table 2. A typical high-risk API source code.

Event	API Source Code
IMEI	Local Telephone Manager.get Imei
Phone number	Local Telephone Manager.get Phonenumber
SMS Center	Get SMS Center
Handled	Value of String
Pid	This M Pid
Install time	Get first Start Time
Sys version	Build VERSION.sdk

2.3. Application Detection Based on Directed Information Flow

This paper proposes a directed information flow method to detect risky applications and reverse query the information leakage path. The data source used was the applica-

tion data obtained from the mobile application market. After reprocessing, it contained 9635 independent applications.

The system permission mechanism is shown in Figure 4. If an application needs to access private data, it needs to apply for the relevant access permissions through the uses-permission tag in manifest.xml to call the API integrated into the system application framework layer to access the system resources and services [18].

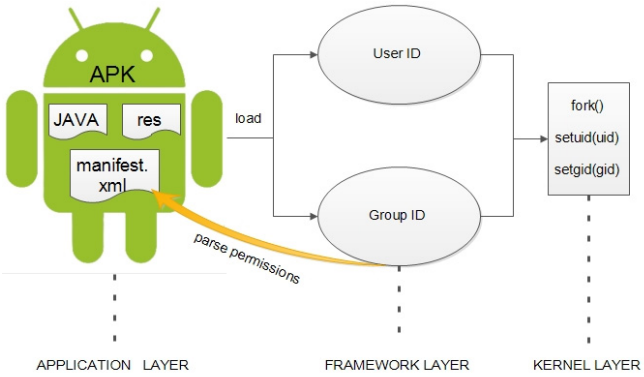

Figure 4. The system permission mechanism.

According to the information flow construction rules and high-risk API list, first, call the reverse tool to analyze the application, then decompile the class.dex file into the Java code file, and analyze the permission statistics results of these applications, as shown in Table 3:

Table 3. The proportion of privacy rights.

Permissions	Application Rate
ACCESS_COARSE_LOCATION	48.7%
ACCESS_FINE_LOCATION	41.5%
GET_TASKS	39.5%
CALL_PHONE	12.1%
READ_SETINGS	10%
READ_ACCOUNTS	10%
GET_ACCOUNTS	9%
SEND_SMS	8%
RECEIVE_SMS	8%

In Table 3, only nine items with the most permissions are listed, of which 48.7% of applications have applied for location access, 41.5% of applications applied for permission to read photo albums, and 39.5% of applications applied for permission to read SMS. More than 98% of all applications were applied for network access.

Next, we utilized getDeviceid() to call the International Mobile Equipment Identity (IMEI) code of the device. If the starting point of the information flow is defined as before this call, the device not only accesses the IMEI number, but also other information after the call. At this time, we tracked the second information tributary, combined with the high-risk source code, and so on [19]. If the last node of the information flow includes information sending and network connection functions, then the software could be considered as a risky application.

This method was used to analyze 100 benign applications and 100 malicious applications, respectively. The benign software was downloaded from the application store, and the malicious software was downloaded from the malicious sample collection website Virus Share. The detection result was that 11 benign applications were marked as risk software, six malicious applications were not successfully identified as risk software, and the rest were correctly identified, so it can be preliminarily considered that the correct

rate of malicious application identification by this method was 94%, and benign applications were correctly identified. The rate was 89%. Applying this method to the collected 9635 independent applications, we found nearly 400 risky applications, and then we verified and analyzed the results to confirm that the detection results were real and effective for the personal data or user account information on the phone.

3. Risk Assessment of Data Leakage Based on Information Entropy and Markov Chain

3.1. Construction of Evaluation Index System

This article incorporated 32 risk indicators into the privacy data leakage evaluation index system, and divided them into five categories according to their attributes: technical level, environmental level, operation management, self-level, and terminal level. The technical level mainly refers to the fact that many applications do not fully consider the security and protection of private data before they are designed. For example, the private data of individual users are not marked and deprived, and the data are often calculated in plaintext. The environmental level mainly refers to the frequent exchange of data in the current network environment and the diversification of privacy. Operation management mainly refers to the data leakage caused by application management personnel such as the malicious leakage of internal personnel, lax advertising review, etc. The user level mainly refers to the lack of privacy awareness of individual users and the simplicity of account passwords. The terminal level mainly refers to the fact that the data do not form a real security closed loop at the terminal, and the privacy protection technology is not perfect, etc. The specific detailed indicators are shown in Table 4.

Table 4. The risk assessment index system of the data leakage of intelligent mobile devices.

Primary Index	Secondary Index	Primary Index	Secondary Index
Technical Level	Intrusion Detection	Operation Management	Advertising Review
	Access Control		Supervision System
	Network Security		Insider Threats
	Anonymous Technology		Third Party Information Collection
	Anomaly Detection		Position Monitoring
	Stain Tracking		Privacy Management
	Identity Authentication	Self Level	Privacy Awareness
	Track Hiding		Intrusion Experience
	Data Sharing		Association Settings
	Data Encryption		Password Settings
Environmental Level	Data Exchange		Permission Setting
	Location Services		Data Identification
	Advertising Attack	Terminal Level	Data Protection
	Protocol Compatibility		Data Control
	Management Regulations		Permission Control
	Privacy Diversity		Event Reminder

Some of the secondary indicators under different primary indicators overlap. For example, the stain tracking at the technical level, the data identification at the own level, and the data control at the terminal level are themselves a risk factor. Therefore, resorting of all of the risk indicators is shown in Figure 5.

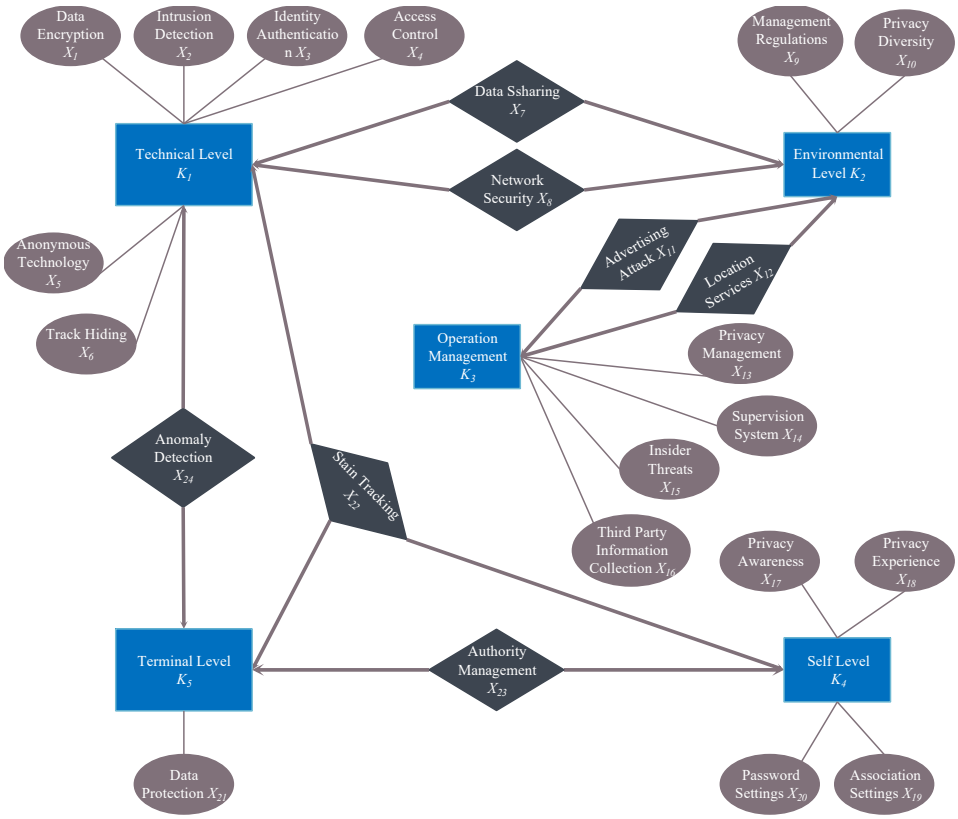

Figure 5. The evaluation index relation diagram.

In Figure 5, the brown ellipse indicates the risk factors belonging to a single category, the black diamond indicates the risk factors belonging to multiple categories, and the blue box indicates the risk categories.

3.2. Risk Assessment of Data Leakage Based on Information Entropy and Markov Chain

Information entropy can be understood as information and entropy. Information refers to all of the information in cyberspace and the object transmitted and processed by communication system [20]. Entropy is a quantity that represents the physical state, which represents the state of an uncertain thing. The greater the quantity of eliminating uncertain factors is introduced, the greater the entropy is. If the certainty is high, there is no need to introduce too many elimination variables, and the entropy is low. Markov chain is a random process algorithm, which means that the state at any time of any random variable completely depends on the state at the previous time, and has nothing to do with the previous state [21]. The characteristics of the Markov chain have the following two aspects. First, the n-step transition is determined by one-step transition, and the n-step transition matrix is the n-th power of the one-step transition matrix. Second, after n-step transitions, the state transition matrix gradually becomes stationary [22].

This article utilized information entropy to solve the characteristics of uncertain transactions, combined with the Markov chain, which could effectively describe the changes of events, and creatively evaluate the risk of the data leakage of intelligent mobile devices. Three smart mobile devices, tablet computer, smart phone, and bracelet, were selected as the research object. Taking the 24 evaluation indices of the above three devices scored by privacy disclosure practitioners in the field of network security for many years as the result, the scores of the questionnaire were all in ten scale, and the expected value and 95% confidence interval of the corresponding indices were obtained. Finally, 237 valid questionnaires were obtained, and the probability value $P(x_i)$ of the corresponding risk factor was obtained as shown in Table 5.

Table 5. The probability of different risk factors for the three mobile devices.

Equipment	Factor	Expect	95% Confidence Interval	Probability	Factor	Expect	95% Confidence Interval	Probability	Factor	Expect	95% Confidence Interval	Probability	Factor	Expect	95% Confidence Interval	Probability
TabletPC	X_1	4.1	3.2–5.6	0.027	X_7	8.0	7.2–9.3	0.052	X_{13}	3.2	2.4–4.0	0.021	X_{19}	8.8	8.3–9.6	0.058
	X_2	4.3	3.5–4.8	0.028	X_8	8.0	7.0–9.2	0.052	X_{14}	4.3	3.5–5.3	0.028	X_{20}	8.2	7.5–9.0	0.054
	X_3	4.6	3.6–5.2	0.030	X_9	3.9	2.8–5.0	0.026	X_{15}	9.1	8.0–9.5	0.060	X_{21}	7.5	6.6–8.4	0.049
	X_4	4.5	2.9–6.0	0.029	X_{10}	4.2	2.7–5.5	0.027	X_{16}	9.2	7.8–9.3	0.060	X_{22}	6.1	5.5–7.0	0.040
	X_5	7.1	5.5–8.7	0.046	X_{11}	3.5	2.8–4.0	0.023	X_{17}	9.2	8.0–9.7	0.060	X_{23}	7.8	5.0–9.3	0.051
	X_6	7.1	6.3–7.5	0.046	X_{12}	4.5	3.5–5.5	0.029	X_{18}	9.6	9.0–10	0.063	X_{24}	5.9	4.8–7.3	0.039
Intelligent mobile phone	X_1	4.2	3.1–5.8	0.031	X_7	5.7	4.5–6.8	0.042	X_{13}	6.8	5.6–7.5	0.050	X_{19}	9.2	6.8–9.8	0.067
	X_2	7.5	6.8–8.5	0.055	X_8	5.8	4.5–7.8	0.042	X_{14}	4.7	3.5–6.4	0.034	X_{20}	3.1	2.0–4.2	0.023
	X_3	4.8	3.0–6.5	0.035	X_9	3.9	3.0–5.2	0.029	X_{15}	4.0	3.3–5.0	0.029	X_{21}	6.5	4.3–8.0	0.048
	X_4	3.9	3.0–6.4	0.029	X_{10}	4.2	2.8–6.4	0.031	X_{16}	8.7	7.3–9.6	0.064	X_{22}	5.5	4.0–6.8	0.040
	X_5	4.1	2.5–6.5	0.030	X_{11}	4.7	2.5–6.2	0.034	X_{17}	8.8	7.5–9.3	0.064	X_{23}	7.8	6.3–8.5	0.057
	X_6	4.0	2.4–7.0	0.029	X_{12}	3.8	2.0–6.7	0.028	X_{18}	9.2	8.5–9.6	0.067	X_{24}	5.9	4.3–7.5	0.043
Bracelet	X_1	2.6	1.5–4.3	0.019	X_7	8.3	7.5–9.6	0.061	X_{13}	8.5	7.5–9.6	0.062	X_{19}	9.2	8.5–9.7	0.067
	X_2	2.7	1.8–4.6	0.020	X_8	4.7	3.2–6.0	0.034	X_{14}	4.7	3.2–7.0	0.034	X_{20}	3.7	2.8–5.6	0.027
	X_3	4.7	3.5–6.0	0.034	X_9	3.5	2.5–4.8	0.026	X_{15}	3.9	2.0–7.5	0.029	X_{21}	4.7	3.0–6.6	0.034
	X_4	3.5	2.5–6.0	0.026	X_{10}	3.7	2.3–5.0	0.027	X_{16}	8.0	5.3–9.7	0.058	X_{22}	5.7	4.0–7.4	0.042
	X_5	2.7	2.0–5.0	0.020	X_{11}	3.6	2.5–5.3	0.026	X_{17}	8.9	7.0–9.9	0.065	X_{23}	8.6	7.6–9.5	0.063
	X_6	8.5	7.2–9.5	0.062	X_{12}	8.7	7.8–9.5	0.064	X_{18}	8.9	7.2–9.9	0.065	X_{24}	4.8	2.6–7.0	0.035

Considering the degree of influence among the risk factors X_i in Table 5, the construction matrix C is as follows:

$$C = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & X_{n3} & X_{nn} \end{bmatrix} \tag{6}$$

In matrix C, the main diagonal element indicates that a risk element occurs alone, while the other two risks exist at the same time. Assuming that a mobile terminal contains only two risk categories K_1 and K_2, K_1 contains X_1, X_2, and X_3, K_2 contains X_3 and X_4, the transfer matrix [23]:

$$P(C) = \begin{bmatrix} P(K_{11}) & P(K_{12}) \\ P(K_{21}) & P(K_{22}) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sum_{i=1}^{3} P(X_i)} P(X_1) + P(X_2) & \frac{1}{\sum_{i=1}^{3} P(X_i)} P(X_3) \\ \frac{1}{\sum_{i=3}^{4} P(X_i)} P(X_3) & \frac{1}{\sum_{i=3}^{4} P(X_i)} P(X_4) \end{bmatrix} \tag{7}$$

Then, for the five primary risk indicators and 24 secondary risk indicators in the above model, the transfer matrix is obtained:

$$P(C) = \begin{bmatrix} P(K_{11}) & P(K_{12}) & \cdots & P(K_{15}) \\ P(K_{21}) & P(K_{22}) & \cdots & P(K_{25}) \\ \vdots & \vdots & \ddots & \vdots \\ P(K_{51}) & P(K_{52}) & \cdots & P(K_{55}) \end{bmatrix} =$$

$$\begin{bmatrix} \frac{1}{\sum_{i=1}^{10} P(X_i)} [P(X_1) + \cdots + P(X_6)] & \frac{1}{\sum_{i=1}^{10} P(X_i)} P(X_7) + P(X_8) & 0 & \frac{1}{\sum_{i=1}^{10} P(X_i)} P(X_{22}) & \frac{1}{\sum_{i=1}^{10} P(X_i)} [P(X_{24}) + P(X_{22})] \\ \frac{1}{\sum_{i=7}^{12} P(X_i)} P(X_7) + P(X_8) & \frac{1}{\sum_{i=7}^{12} P(X_i)} P(X_9) + P(X_{10}) & \frac{1}{\sum_{i=7}^{12} P(X_i)} P(X_{11}) + P(X_{12}) & 0 & 0 \\ 0 & \frac{1}{\sum_{i=11}^{16} P(X_i)} P(X_{11}) + P(X_{12}) & \frac{1}{\sum_{i=11}^{16} P(X_i)} [P(X_{13}) + \cdots + P(X_{16})] & 0 & 0 \\ \frac{1}{\sum_{i=17}^{23} P(X_i)} P(X_{22}) & 0 & 0 & \frac{1}{\sum_{i=17}^{23} P(X_i)} [P(X_{17}) + \cdots + P(X_{19})] & \frac{1}{\sum_{i=17}^{23} P(X_i)} [P(X_{23}) + P(X_{22})] \\ \frac{1}{\sum_{i=21}^{24} P(X_i)} [P(X_{22}) + P(X_{24})] & 0 & 0 & \frac{1}{\sum_{i=21}^{24} P(X_i)} [P(X_{23}) + P(X_{22})] & \frac{1}{\sum_{i=21}^{24} P(X_i)} P(X_{21}) \end{bmatrix} \tag{8}$$

4. Discussion

Obtain the result: $P^{com}(C)$, $P^{tel}(C)$, and $P^{bra}(C)$:

$$P^{com}(C) = \begin{bmatrix} 0.567 & 0.287 & 0 & 0.110 & 0.218 \\ 0.498 & 0.254 & 0.249 & 0 & 0 \\ 0 & 0.235 & 0.765 & 0 & 0 \\ 0.123 & 0 & 0 & 0.555 & 0.279 \\ 0.441 & 0 & 0 & 0.508 & 0.274 \end{bmatrix}$$

$$P^{tel}(C) = \begin{bmatrix} 0.592 & 0.238 & 0 & 0.113 & 0.235 \\ 0.408 & 0.291 & 0.301 & 0 & 0 \\ 0 & 0.259 & 0.741 & 0 & 0 \\ 0.125 & 0 & 0 & 0.623 & 0.305 \\ 0.441 & 0 & 0 & 0.516 & 0.255 \end{bmatrix}$$

$$P^{bra}(C) = \begin{bmatrix} 0.550 & 0.289 & 0 & 0.128 & 0.234 \\ 0.399 & 0.227 & 0.378 & 0 & 0 \\ 0 & 0.329 & 0.670 & 0 & 0 \\ 0.128 & 0 & 0 & 0.599 & 0.319 \\ 0.443 & 0 & 0 & 0.603 & 0.195 \end{bmatrix}$$

$P^{com}(C)$, $P^{tel}(C)$, and $P^{bra}(C)$ are the risk factor transfer matrices of the tablet computer, mobile phone, and wristband, respectively.

The process of finding the steady-state probability of various risks is to find the eigenvector of the state transition matrix. Because the state transition matrix is full rank, the solution vector is unique, and the elements in the solution vector are the steady-state probability value of each risk category.

The steady-state probability $\hat{P}(K_i)$ of K_i and matrix $P(C)$ satisfy:

$$\begin{cases} \hat{P}(K_1) = P(K_{11})\hat{P}(K_1) + P(K_{12})\hat{P}(K_2) + \cdots + P(K_{1m})\hat{P}(K_m) \\ \hat{P}(K_2) = P(K_{21})\hat{P}(K_1) + P(K_{22})\hat{P}(K_2) + \cdots + P(K_{2m})\hat{P}(K_m) \\ \vdots \\ \hat{P}(K_5) = P(K_{51})\hat{P}(K_1) + P(K_{52})\hat{P}(K_2) + \cdots + P(K_{5m})\hat{P}(K_m) \\ 1 = \hat{P}(K_1) + \hat{P}(K_2) + \cdots + \hat{P}(K_5) \end{cases} \tag{9}$$

The steady-state probability values of the three devices are calculated as follows:

$$P^{com}(K_i) = (0.276, 0.175, 0.195, 0.103, 0.251)^T$$

$$P^{tel}(K_i) = (0.236, 0.205, 0.139, 0.176, 0.244)^T$$

$$P^{bra}(K_i) = (0.277, 0.196, 0.105, 0.162, 0.260)^T$$

Then, bring $\hat{P}(K_i)$ into the information entropy formula to obtain:

$$H = -\sum_{i=1}^{5} \hat{P}(K_i) \log_2 \hat{P}(K_i) \tag{10}$$

Normalize H to obtain the entropy values of the three devices $H^{com} = 2.251$, $H^{tel} = 2.294$, $H^{bra} = 2.246$. The risk situation of different categories of each device is shown in Figure 6.

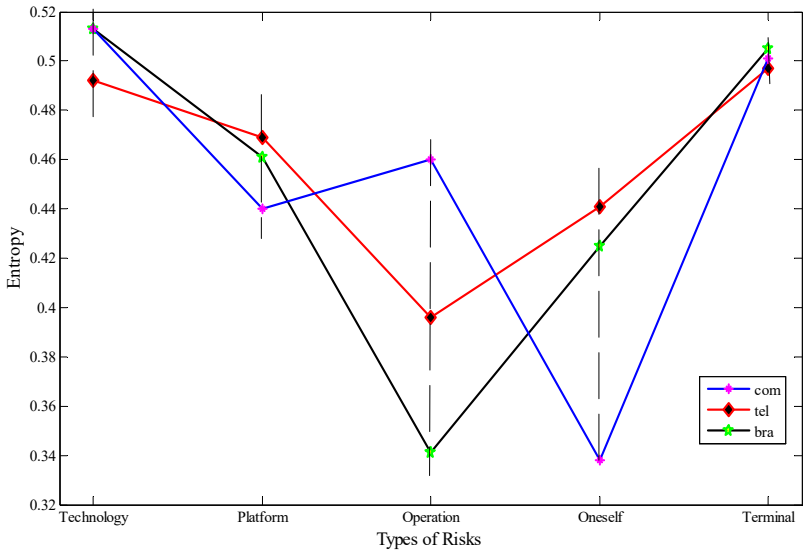

Figure 6. The risk assessment results of the different types of equipment.

In Figure 6, the blue, red, and black lines, respectively, represent the entropy under each risk category of the tablet computer, smartphone, and bracelet. The ordinate represents the entropy value, and the abscissa represents the primary index of the dataset. First of all, tablet computers have the largest entropy at the technical level, followed by the terminal level, followed by the operation risk, platform, and self. The greater the entropy and

the higher the uncertainty, the greater the possibility of information disclosure. Tablet computers are prone to privacy disclosure at the technical and terminal levels. Similarly, smartphones and bracelets are prone to information leakage at the technical and terminal levels. Conversely, smartphones and bracelets are stable at the level of operation risk, which is not easy to cause information leakage, while tablets are stable at their own level. Overall information leakage risk: according to the entropy obtained by the above three devices, the overall information leakage risk of the three devices is almost the same. From a micro perspective, $P^{tel}(C) > P^{com}(C) > P^{bra}(C)$, it shows that the information leakage risk is in the order of the smartphone, tablet, and bracelet from large to small.

5. Conclusions

This paper mainly studied malicious application detection and information leakage risk assessments. First, this paper used the directed information flow algorithm, high-risk API source code, and reverse tools to detect malicious software applications and hardware systems. Second, the risk assessment of information leakage events of intelligent mobile devices was carried out. The research objects were tablet computers, smartphones, and bracelets in smart mobile devices. Generally speaking, there was little difference in the entropy of data risk assessment among the three, but there were differences in the different types of risks. According to the expectation of the tenth system, the risk of the three was low, and there was a certain degree of privacy disclosure. In terms of data operation and management, the risk value of the computer was higher than that of the mobile phone and bracelet. However, in terms of its own risk, the mobile phone was higher than that of the bracelet and computer, indicating that the operation environment of computers should be strengthened. The mobile phone and bracelet need to consolidate the firewall to reduce their own risk in the process of developing software and hardware. At the level of technology, platform, and terminal, entropy is high and the difference is small. In order to provide a more assured and pleasant network experience to network users, operators should strengthen the control and optimization of the network environment and network platform, identify and encrypt the users' private data, and accelerate development. The hardware-supported isolation environment performs safe and efficient plaintext calculations on key codes and data, and hides the calculation mode to prevent data holders from inferring private data, strengthens identity authentication and confidentiality agreements, and ensures that user privacy data are not leaked [24]. The boundary of an app's collection of personal privacy should be based on whether the user needs it or not, and more consideration should be given to the relevant rights and interests of the user [25]. Through this model research, it also reflects the disadvantages of the current intelligent mobile devices, and provides constructive guidance for intelligent device manufacturers and the operators' network construction.

Author Contributions: Conceptualization, X.Y.; Data curation, X.Y.; Formal analysis, X.Y.; Funding acquisition, Y.L.; Investigation, X.Y.; Methodology, X.Y. and Y.L.; Project administration, Y.L.; Software, X.Y. and J.X.; Supervision, X.Y.; Validation, X.Y.; Visualization, X.Y.; Writing—original draft, X.Y.; Writing—review & editing, X.Y. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 61972334).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (No. 61972334).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, X.; Chen, H. A review of high-dimensional data publishing research on differential privacy. *CAAI Trans. Intell. Syst.* **2021**, *16*, 989–998. [CrossRef]
2. Zhang, T. *Research on Risk Factors and Risk Assessment Methods of User Privacy Disclosure in Mobile Commerce*; Yunnan University of Finance and Economics: Kunming, China, 2021.
3. Guo, Y.; Duan, Q.S.; Wang, X.W. An Empirical Study on Privacy Information Disclosure Behaviour of Mobile Learning Users. *J. Mod. Inf.* **2018**, *38*, 98–117.
4. Xiong, J. Research on privacy information disclosure behavior and protection of mobile commerce users—From the perspective of evolutionary game theory. *Fortume Times* **2018**, *2018*, 63–64.
5. Wang, K. *Evidence Theory Based Evaluating and Controlling Mobile Commerce Transactions Risk*; Huazhong University of Science and Technology: Wuhan, China, 2009.
6. Zhao, Z.H. *An Empirical Study on the Determinants of Intentions to Use Mobile SNS Applications—Take "WeiXin" for Example*; Shandong University: Jinan, China, 2014.
7. Li, Y.H.; Liang, L.T.; Liu, B.L. An Empirical Study on Privacy Beliefs and Information Disclosure Willingness of Mobile Social Users. *Inf. Theory Pract.* **2016**, *39*, 76–81.
8. Xu, J.L.; Qiao, Z.; Wang, X.Q.; Li, F. Research and Application of Privacy Information Detection and Protection Technology for Mobile Internet Users. *Telecom Eng. Tech. Stand.* **2019**, *2019*, 12–22.
9. Mark, F.; Alexander, B. Do privacy concerns matter for Millennials? Results from an empirical analysis of Location-Based Services adoption in Germany. *Comput. Hum. Behav.* **2015**, *53*, 344–353.
10. Jia, J. *The Research of Personal Privacy Information Security in the Era of Big Date*; Neimenggu University: Huhehaote, China, 2018.
11. Wu, J.Z.; Wu, Y.J.; Wu, Z.F.; Yang, M.T.; Luo, T.Y.; Wang, Y.J. An Android privacy leakage malicious application detection approach based on directed information flow. *J. Univ. Chin. Acad. Sci.* **2015**, *32*, 807–815.
12. Jin, X.Q.; Lu, J.Q.; Li, L.C. Design of network anomaly detection and intrusion prevention system based on information entropy. *Electron. Des. Eng.* **2021**, *29*, 152–156.
13. Zhang, Z.G.; Wang, X.J.; Li, G.; Yue, S.M. The Generation Method of Network Defense Strategy Combining with Attack Graph and Game Model. *Netinfo Secur.* **2021**, *21*, 1–9.
14. Song, X.M. *Research on Covert Channel Identification Methods Based on Semantic Information Flow*; Jiangsu University: Zhenjiang, China, 2017.
15. Yang, T. *Research on Detection Methods of Communication Privacy Leakage of Smart Home System*; Hebei University of Science and Technology: Shijiazhuang, China, 2020.
16. Pan, C.J. *Research on Private Information Disclosure Detection Method of Composite Services*; Xidian University: Xi'an, China, 2019.
17. Russo, A.; Lax, G.; Dromard, B.; Mezred, M. A System to Access Online Services with Minimal Personal Information Disclosure. *Inf. Syst. Front.* **2021**. [CrossRef]
18. Sun, C.G.; Zhu, W.Z.; Li, W.F.; He, X. A method for detecting privacy data leakage in Android application. *J. Zhengzhou Univ. Sci. Ed.* **2019**, *52*, 68–74.
19. Peng, Y.C. Consideration and analysis of public information disclosure and personal information protection in epidemic response. *Chin. J. Gen. Pract.* **2021**, *19*, 1760–1763.
20. Yang, A.; Liu, H.; Chen, Y.; Zhang, C.; Yang, K. Digital video intrusion intelligent detection method based on narrowband Internet of Things and its application. *Image Vis. Comput.* **2020**, *97*, 130914. [CrossRef]
21. Chen, W.; Lv, W.Y.; Li, S.Q.; Dai, J.; Deng, X. Estimation and Comparison of Two Markov Chain State Transition Probability Matrices. *J. Chongqing Univ. Technol. Nat. Sci.* **2021**, *35*, 217–223.
22. Jiang, L.; Liu, J.Y.; Wei, Z.B.; Gong, H.; Lei, C.; Li, C.X. Running State and Its Risk Evaluation of Transmission Line Based on Markov Chain Model. *Autom. Electr. Power Syst.* **2015**, *39*, 51–58.
23. Song, L.J.; Xu, Z.Y. Assessment of power customer credit risk based on set pair analysis and Markov chain model. *Electr. Power Autom. Equip.* **2009**, *29*, 37–40.
24. Pettai, M.; Laud, P. Combining differential privacy and secure multiparty computationl. In Proceedings of the 31st Annual Computer Security Applications Conference, Los Angeles, CA, USA, 7–11 December 2015; pp. 421–430.
25. Zhu, X.X.; Liu, X.Y.; Xiong, Q.Q. Research on the impact of App permissions on user privacy. *Wirel. Internet Technol.* **2021**, *18*, 13–18, 41.