THREAT HUNTER RESEARCH

2022 Data Asset Exposure Report

An analysis of 2022 data-exposure patterns, underground trading activity and the role of third-party governance.

Original publication: 2023-03-02
Research team: Threat Hunter Research Team
Source report: 28 pages

Original report text

This text version is reconstructed based on the 28-page original PDF, retaining the report narrative, chapters and research scope; the cover, repeated table of contents and purely decorative pages are not repeated. Localized figures are placed in context throughout this web edition, with the original PDF retained for reference.

Preface

While big data and the Internet bring unlimited development vitality to enterprises, they also hide huge data security risks. With the in-depth development of digital transformation, the frequency, scale and cost of data breaches have also increased compared with previous years, causing continuous and difficult to eliminate impacts on many enterprises.

In February 2022, the resume data of mainstream domestic recruitment platforms was suspected of being crawled, involving more than 210 million pieces of personal information;

In July 2022, more than 220 million user IDs and phone numbers on a domestic video sharing platform were suspected to be sold;

In November 2022, a major international social media company was fined US$265 million after crawlers scraped its platform and exposed the data of 533 million users;

...

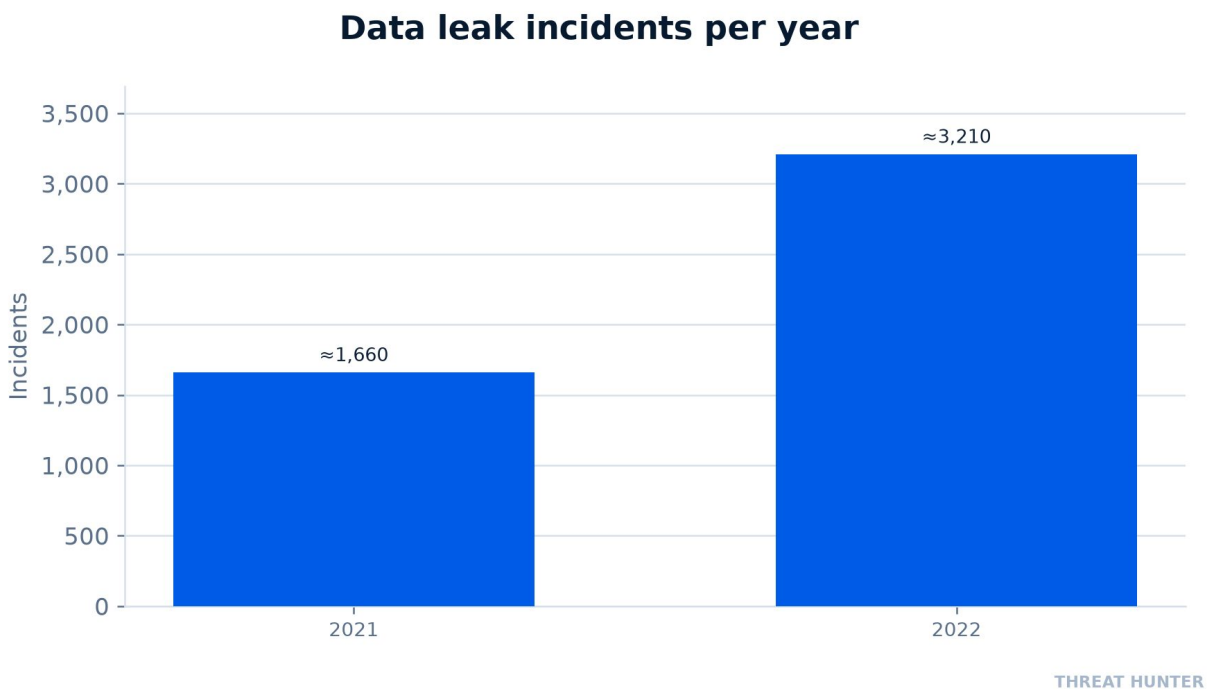
Threat Hunter's "2022 Data Asset Leakage Analysis Report" is based on data leakage incidents captured in the past year, combined with the current status of data security, to present a multi-dimensional panoramic view of domestic data leakage situations in 2022.

Data breach risk profile

1. Overview of data leakage risks in 2022

1. A total of more than 3,200 data leakage incidents were captured in 2022, nearly an increase from 2021

The Threat Hunter intelligence platform discovered and verified more than 3,200 valid data breaches in 2022, nearly doubling compared with 2021. The number of data breaches in the second half of the year increased significantly compared with the first half.



Data leak incidents captured, 2021 vs 2022: the 2022 total nearly doubled to more than 3,200

Source: Threat Hunter original report, page 5. Redrawn in English; values read from the original chart and approximate.

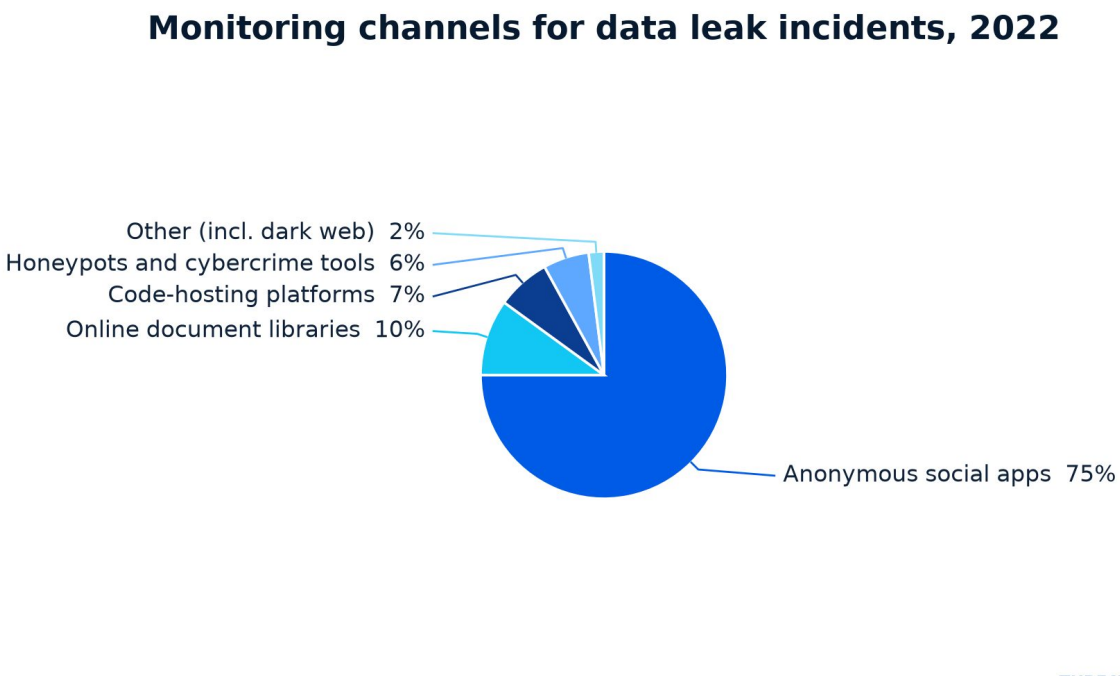


Monthly data leak incidents in 2022: a peak of about 357 in October

Source: Threat Hunter original report, page 5. Redrawn in English; values read from the original chart and approximate.

2. Data leakage channels come from a wide range of sources, with anonymous social software accounting for over 75%

The following is the proportion of data breaches discovered by Threat Hunter intelligence researchers from various channels in 2022. More than 75% of data breaches were discovered through anonymous social software, led by one encrypted messaging app.



Monitoring channels for data leak incidents in 2022: anonymous social apps account for 75%

Source: Threat Hunter original report, page 6. Redrawn in English from the original figure.

In 2022, the main channels through which the Threat Hunter intelligence platform captures data breaches are as follows:

Anonymous social software: Anonymous social and encrypted messaging apps are used by threat actors as a platform for data transactions and are currently the main channel for capturing data leakage incidents;

Online cloud drives and document libraries: Some corporate employees use online cloud drives and document-sharing libraries as tools for file storage and document sharing. Due to lack of security awareness and other reasons, some files or documents containing sensitive information may be mistransmitted and can be publicly accessed by anyone, eventually leading to data leakage;

Code hosting platform: Some developers will upload the code they write to public code-hosting platforms, which may expose sensitive information such as backend addresses, accounts, and passwords.

In 2017, the source code of a carmaker's connected-car application was published on a public code-hosting platform and was maliciously exploited, leaking the emails of 300,000 customers. It was not discovered and changed until October 2022, five years later;

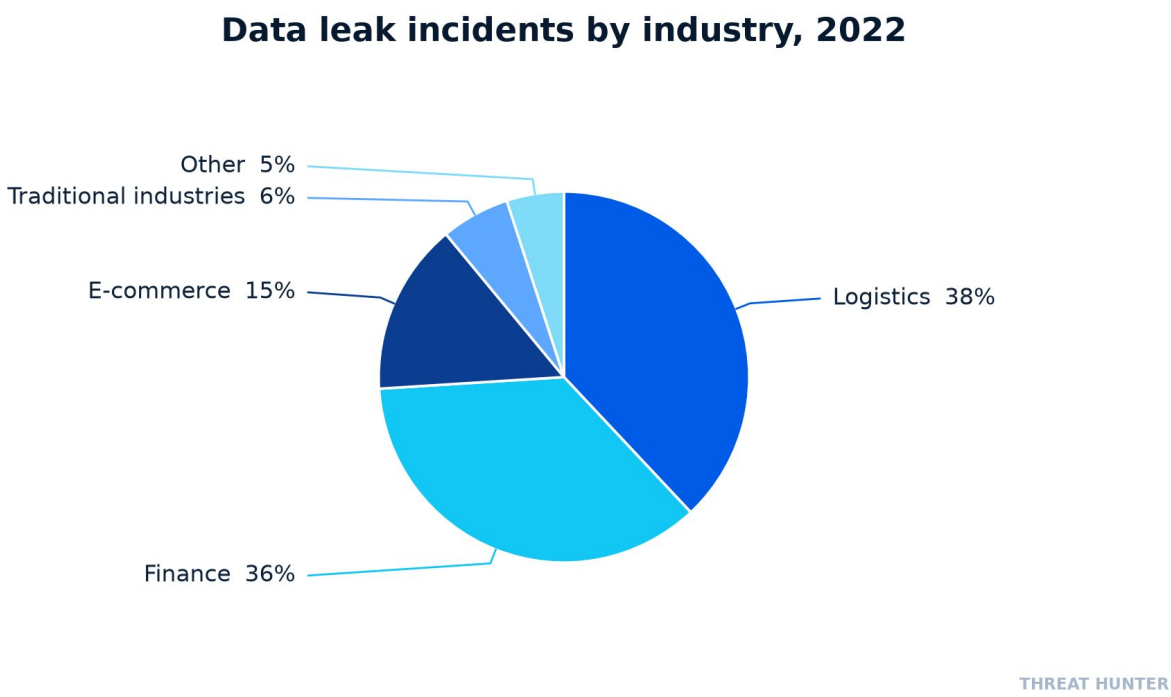
Honeypot + threat actors tool: Threat Hunter's unique honeypot technology and threat actors tool analysis technology capture network cybercriminal groups attack traffic in real time. By analyzing these attack traffic, a large number of data-stealing attacks can be discovered, becoming one of the important channels for discovering data leakage incidents.

Darknet: A darknet network built based on anonymous communication technologies such as Tor and I2P. There are various underground trading markets and forums. Many major data leaks in history were first exposed on the darknet;

In addition, it also includes threat actors forums, public media and other channels represented by Raid Forums, Breach Forums, etc.

3. In the distribution of data breach industries in 2022, finance, logistics, and e-commerce industries occupy the top three

From the perspective of industry distribution, among the data breaches captured in 2022, the financial industry, logistics industry, and e-commerce industry accounted for a large proportion.



Data leak incidents by industry in 2022: logistics 38%, finance 36%, e-commerce 15%

Source: Threat Hunter original report, page 7. Redrawn in English from the original figure.

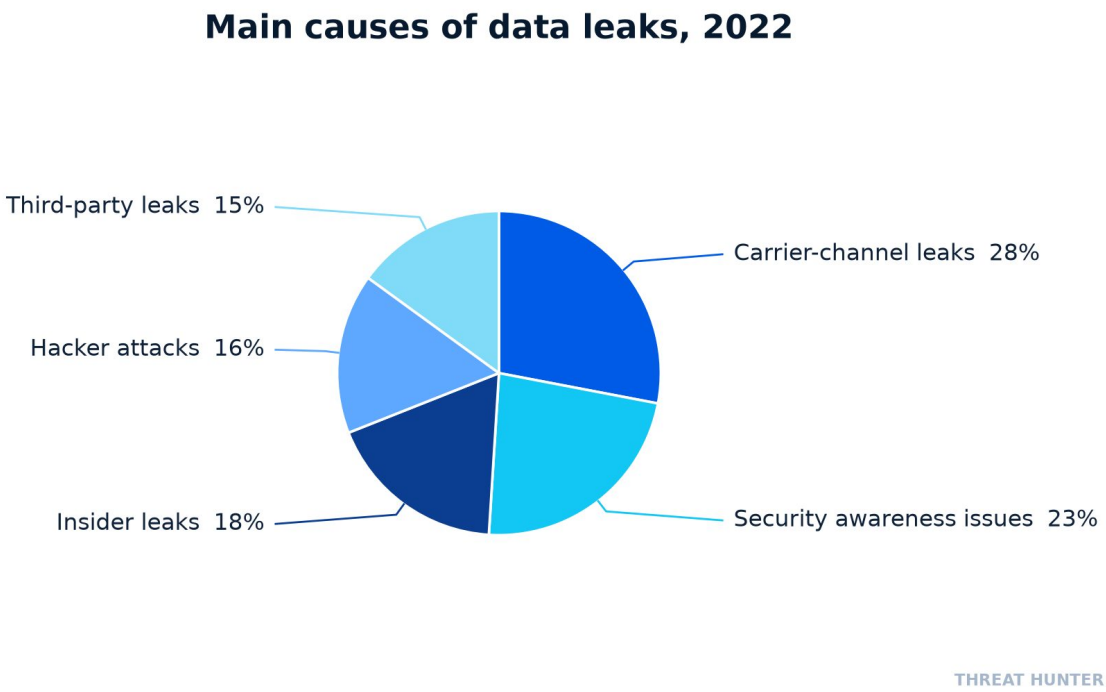
The financial industry has become the main target of attacks by threat actors because it involves a large amount of high-value user data and is close to the transaction link. Data breaches in the financial industry are mainly reflected in the reselling of customer information by lending, banking, securities, insurance and other companies, which are usually used by downstream threat actor groups for targeted marketing and fraud.

With the rapid development of online shopping, a large amount of shopping orders and logistics information have been generated. Due to their large exposure, these shopping orders and logistics information have gradually become the key targets of threat actors and are also used for marketing or fraud.

4. Among the main causes of data leakage in 2022, operator channel leakage accounts for the first place

Among the data leakage incidents captured by Threat Hunter in 2022, some incidents can be located through source analysis to locate the cause of the leakage. According to statistical analysis

by Threat Hunter intelligence experts, the main causes and proportions of data breaches in 2022 are as follows:



Main causes of data leaks in 2022: carrier-channel leaks rank first at 28%

Source: Threat Hunter original report, page 8. Redrawn in English from the original figure.

Operator channel: threat actors obtain information such as access data of specified web pages, installation data of specified applications, receiving and sending data of specified text messages, etc. through illegal agents or operator insiders;

Security awareness issues: During the work process, internal employees of the company unconsciously upload important files, documents, codes, etc. within the company to public network environments such as network disk libraries and code hosting platforms;

Insider leaks: Internal employees of the company, driven by interests, use methods such as data export and manual photography to obtain sensitive customer information and then sell it;

Hacker attack: Hackers use crawlers, scanning, penetration and other methods to attack the network assets and systems of enterprises, find security vulnerabilities and then steal data on a large scale;

As the level of network security in various industries continues to improve, it is becoming more and more difficult for hackers to invade and obtain backend permissions. Therefore, many hackers target API interfaces with security flaws and steal data through data traversal attacks and other methods.

Third-party leakage: A third party that has a cooperative relationship with an enterprise has permission to access some sensitive data of the enterprise. Due to problems such as irregular data permission management, these sensitive data are leaked to threat actors through the third party.

threat-actor data transaction analysis

2. Analysis of underground black market data transactions in 2022

1. Encrypted messaging apps replace the darknet trading market and becomes the most important platform for data trading

The "Ancient Tea Horse Road", once one of the largest Chinese darknet trading markets, was destroyed by the police at the end of 2021. Subsequently, darknet trading markets such as "Chang'an Evernight City" and "Free Country" developed.

There is not a small amount of data sales information released by the darknet trading market in 2022, but after analysis and verification, the vast majority of the data released by the darknet trading market is historical leaked data, false data or leakage data, and the proportion of new real leaked data is smaller.

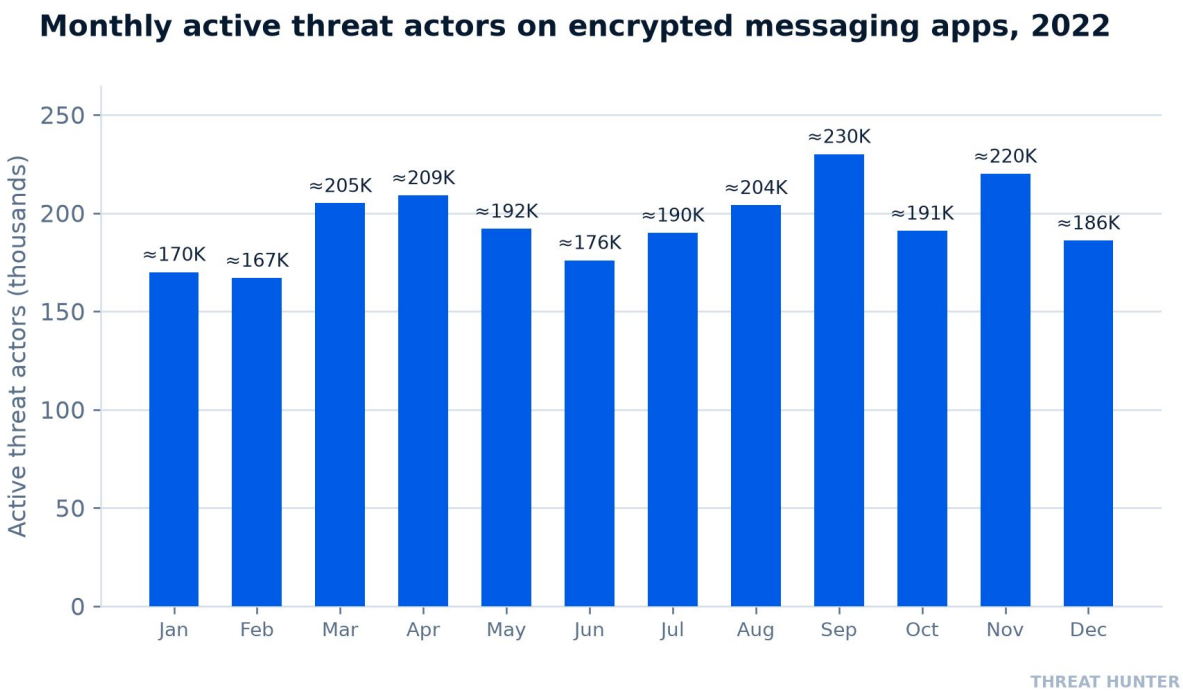
After research and analysis, the main reasons are as follows:

1. Due to the multiple rounds of police crackdowns on darknet markets, many data sellers no longer regard it as the preferred platform;

2. Due to the inability to withdraw cash and administrators running away in some darknet trading markets, sellers’ trust in the platform has been greatly reduced.

In comparison, anonymous encrypted messaging apps have gradually replaced the darknet trading market and become the most important platform for threat actors to illegally trade data. In 2022, Threat Hunter monitored more than 1.46 million active threat actors on encrypted messaging apps, with an average of more than 195,000 active per month.

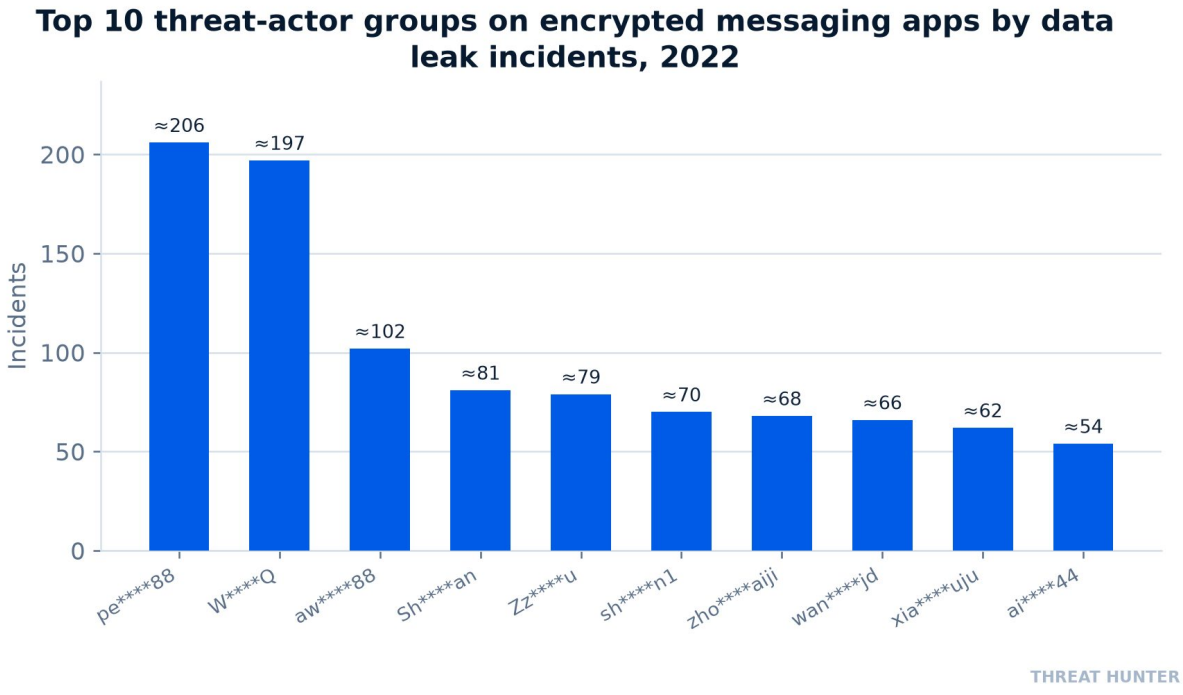
After research and investigation, it was found that among the threat-actor groups that have been active on these apps for a long time, the top ten threat-actor groups ranked in the number of data transactions in 2022 have conducted more than 50 data transactions, and the top two threat-actor groups have conducted about 200 data transactions in total.



Monthly active threat actors on encrypted messaging apps in 2022: about 195,000 a month on average

Source: Threat Hunter original report, page 12. Redrawn in English; values read from the original chart and approximate.

These threat actors have created multiple group chats, channels, and bots on these apps. Once they have new data, they will publish sales information on these channels to attract more buyers. On February 12, the suspected leak of 4.5 billion address information attracted widespread attention. The gang first released relevant data information on its messaging-app channel.



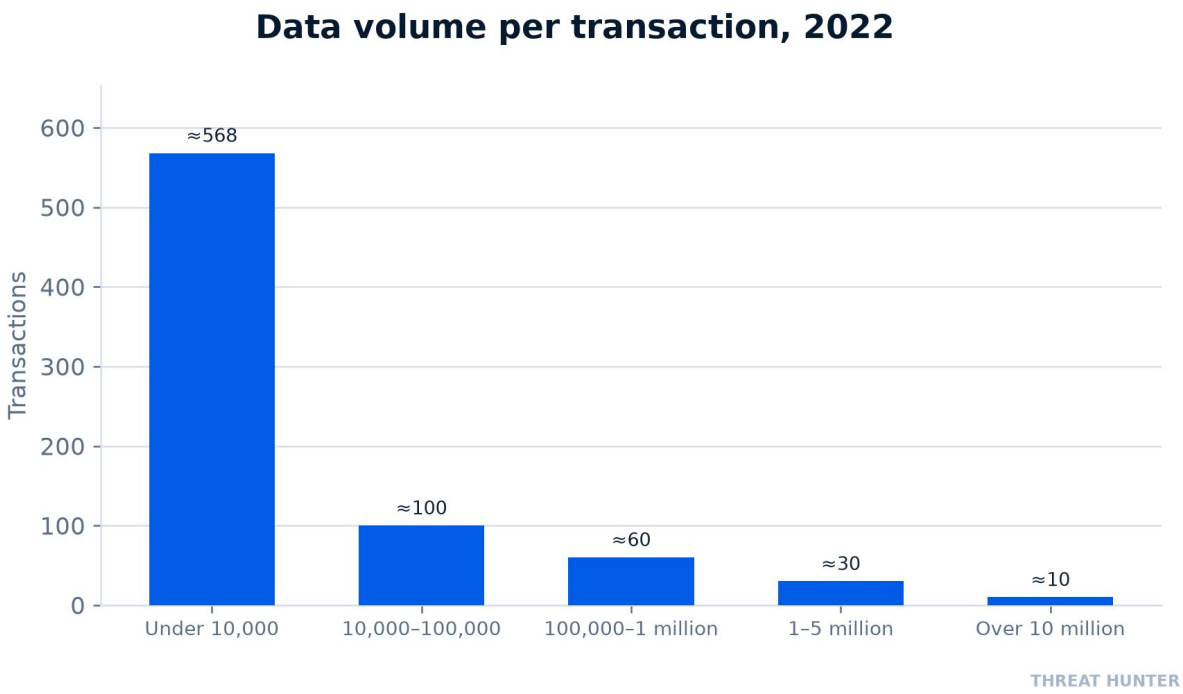
Data leak incidents of the top 10 threat-actor groups trading on encrypted messaging apps, 2022 (names masked as in the original)

Source: Threat Hunter original report, page 12. Redrawn in English; values read from the original chart and approximate.

2. Small-scale real-time data becomes the mainstream of data transactions

In 2022, among the underground black market data transactions captured by the Threat Hunter intelligence platform, about 71% of the transactions disclosed the amount of data sold, of which "small-scale real-time data" with transaction data magnitude below 1W exceeded 73%, becoming the mainstream of data transactions.

According to research and analysis by Threat Hunter, the reasons why small-scale data transactions have become mainstream are as follows:



Magnitude of data sold per transaction in 2022: deals under 10,000 records dominate

Source: Threat Hunter original report, page 13. Redrawn in English; values read from the original chart and approximate.

1. In some data leakage channels, the data that sellers can obtain at one time is often less, and buyers often need to place orders in advance;

2. Although these data are not large in magnitude, they are basically real-time or quasi-real-time data with good validity, so they are more favoured by buyers;

3. The unit price of real-time data is often relatively high, and the cost of purchasing a large amount of real-time data at one time is too high.

In comparison, due to the large volume and high cost of large-scale data transactions, few buyers can afford it, and there may be a lot of historical data in them, so the data value is limited, so the proportion of large-scale data transactions is greatly reduced.

3. Frequent incidents of trading using a combination of “historical data + new data”

In order to improve the effect of targeted marketing or fraud, downstream threat actors often need to master as much victim information as possible, so sellers will also conduct transactions by combining data, the most typical of which is "historical data" combined with "new data."

For historical data, many data sellers have mastered massive historical leak libraries, and a large amount of sensitive personal information such as names, phone numbers, ID cards, and addresses can be extracted from historical data.

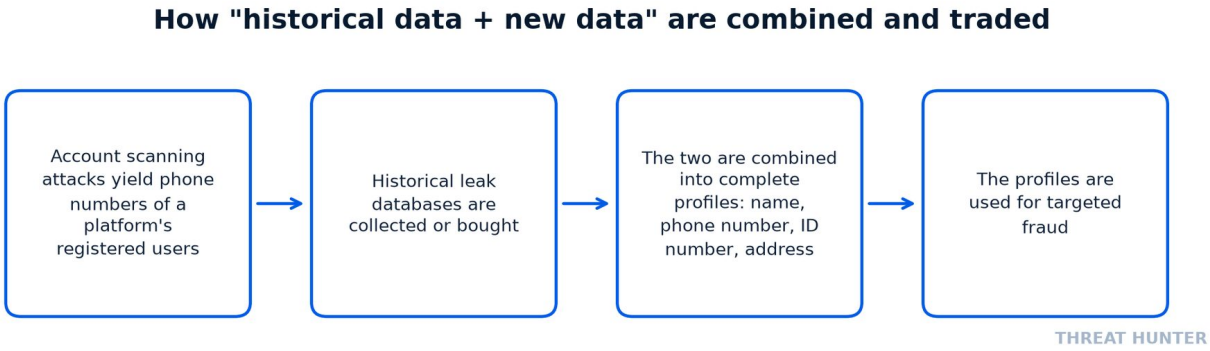
For new data, cybercriminal groups used API interfaces with security flaws in some financial, e-commerce and other platforms to screen out the phone numbers of registered users of the platform through batch number scanning attacks.

Using the phone number as a link and combining historical data with new data, a comprehensive data can be obtained, including the phone number, name, ID card, address and other information of registered users on a certain platform.

In 2022, the Threat Hunter intelligence platform captured a large number of account scanning attacks through honeypot technology, covering multiple platforms. Among the phone numbers passed in by the attacker, many match those in the historical leak database.

Data leakage case

3. Data breach cases worthy of attention in 2022



How "historical data + new data" are combined and traded

Source: Threat Hunter original report, page 14. Redrawn in English from the original figure.

1. A total of more than 1,100 data leakage incidents from financial institutions have been captured, involving more than 300 companies.

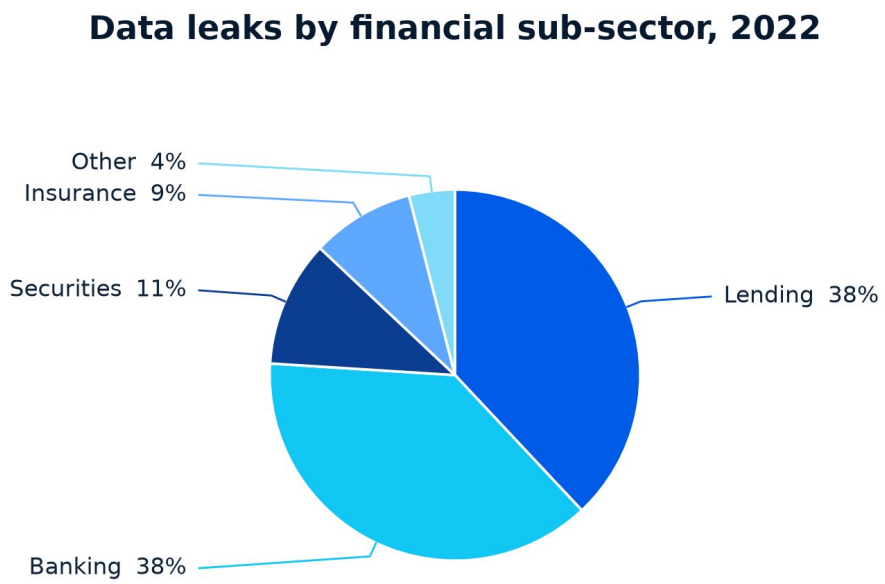
Financial institutions have accumulated a large amount of high-value data in the process of conducting business. These high-value data have always been "sweet" in the eyes of threat actors. In Threat Hunter data leakage monitoring intelligence, financial institutions' data leakage incidents often rank at the top.

In 2022, Threat Hunter captured more than 1,100 data breaches related to financial institutions, involving more than 300 companies, covering multiple segments such as banking, securities, insurance, and lending.

In November 2022, the Threat Hunter intelligence platform monitored more than 80,000 pieces of customer information data leaked by a domestic bank. The leaked data format includes customer names, ID numbers, dates of birth, loan status, phone numbers, whether there is a mortgage, marital status, education and other sensitive fields.

This type of data is often used by downstream threat actor groups for targeted marketing and fraud. For example, by obtaining customer loan information, contacting customers through text messages, phone numbers, etc., for targeted marketing of related loan platforms, or by pretending to be a bank staff member to report customer-related information to induce users to make transfers or other illegal operations.

Currently, financial institutions are relatively mature in data security construction. However, driven by digital interactions and interests, customer information is still stolen by threat actors through operator channels, third-party tools/platforms, etc., bringing unprecedented risks and challenges to the "strictly regulated, high-standard" financial industry.

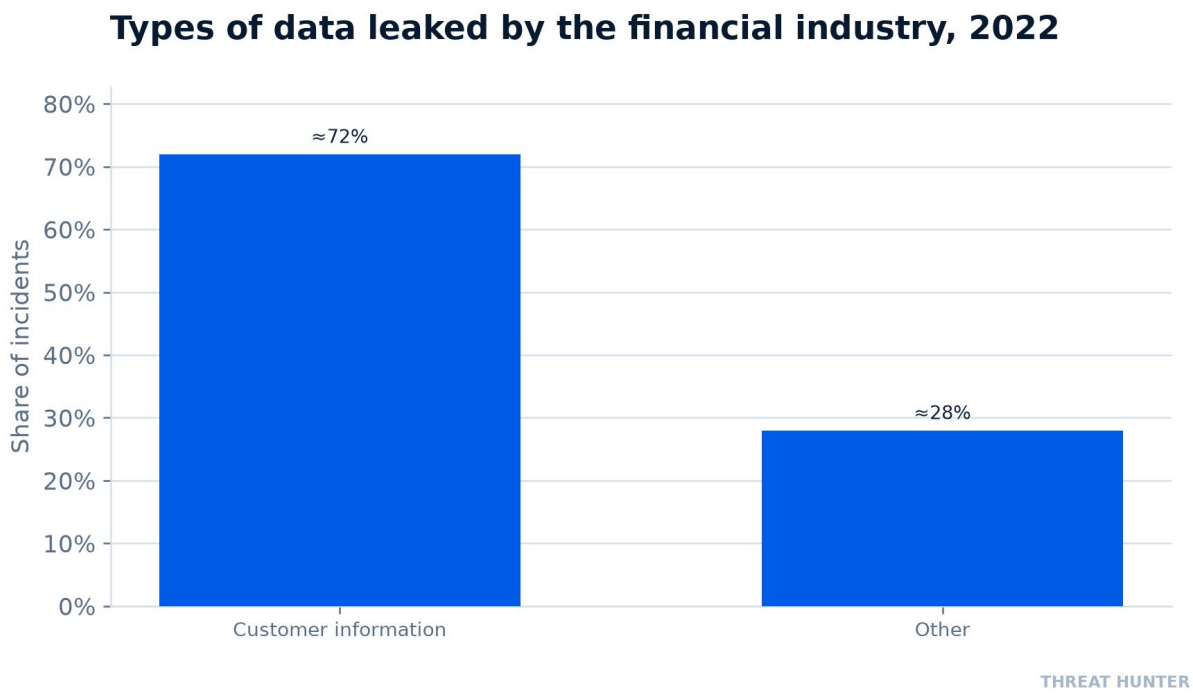


Data leaks by financial sub-sector in 2022: lending and banking account for 38% each

Source: Threat Hunter original report, page 16. Redrawn in English from the original figure.

Data leakage incidents occur frequently in financial institutions caused by operator channels and third-party tools/platforms. From July to August 2022, the Threat Hunter intelligence platform captured multiple data leaks of text message content of customers of domestic lending platforms. After research, it was found that the main cause of the leaks was operator channel leaks. The leaked data format mainly includes sensitive fields such as user phone numbers, text message content, time of receiving text messages, and names.

In 2022, the Threat Hunter Karma intelligence platform monitored that a third-party collection company developed automated "collection assistance tools" through external tool authors, involving multiple financial institutions. After research and analysis, it was found that the "collection assistance tool" can bypass the collection system, gain access to user data by cracking and forging "API interface for pulling data" requests, and download customer data in batches (including borrower's name, ID card, loan amount, address, etc.) without being affected by system access rights and operation restrictions, posing a greater risk of data leakage to financial institutions.

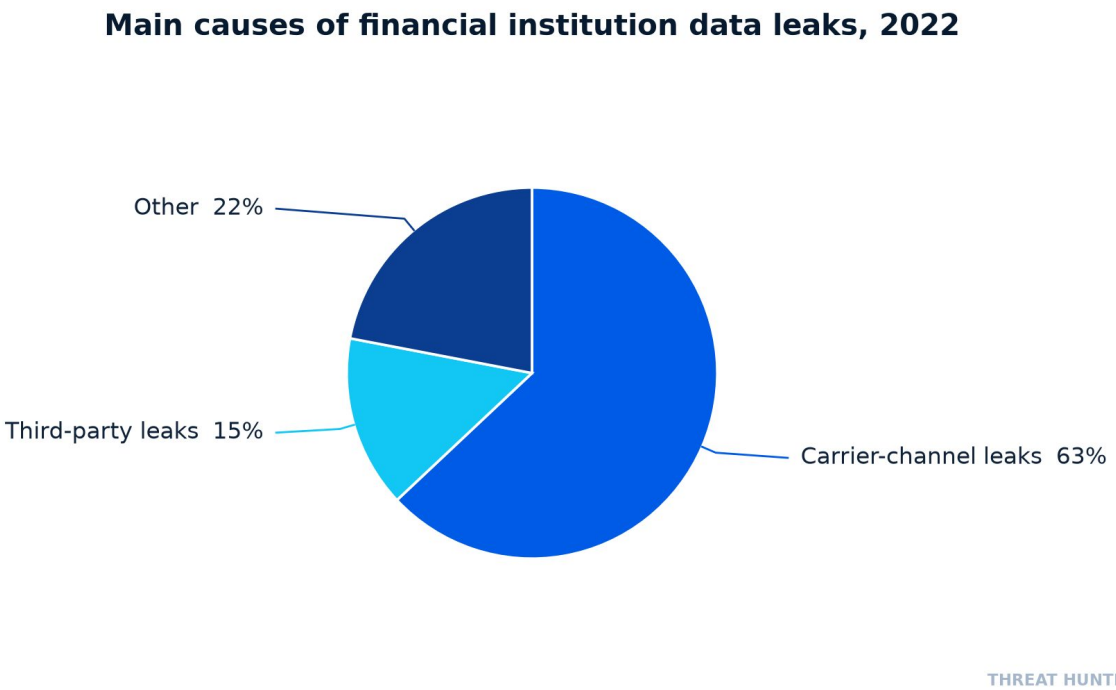


Types of data leaked by the financial industry in 2022: customer information accounts for about 72%

Source: Threat Hunter original report, page 17. Redrawn in English; values read from the original chart and approximate.

2. A total of more than 1,200 data leakage incidents in the logistics industry have been captured, with the highest leakage magnitude exceeding 100

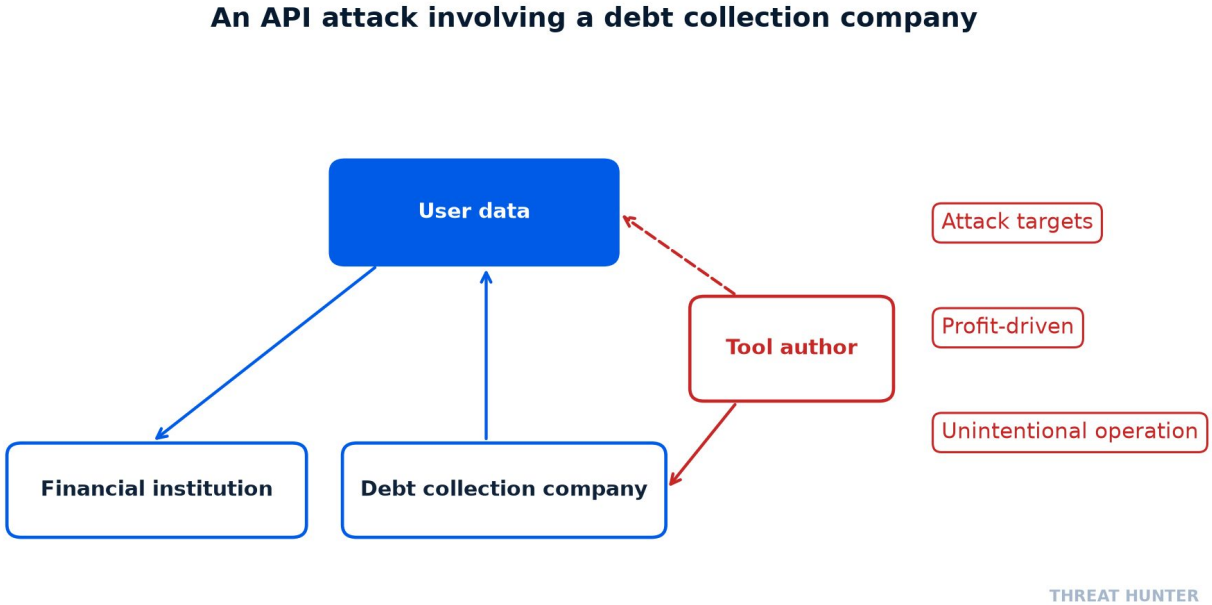
In 2022, Threat Hunter has captured a total of more than 1,200 data leaks related to the logistics industry, involving 40 logistics and express companies. The maximum leakage magnitude exceeded 1 million. On average, more than 10,000 pieces of user information are traded in the underground black market every day, and the price is basically maintained at 3-4 yuan per piece. The main leakage is caused by express delivery orders. The leaked information includes sensitive fields such as the recipient's name, phone number, product information, express delivery order number, and time.



Main causes of financial institution data leaks in 2022: carrier-channel leaks account for 63%

Source: Threat Hunter original report, page 18. Redrawn in English from the original figure.

After investigation and analysis, the main reasons for the leakage of express delivery orders include:



An API attack involving a debt collection company in the financial industry

Source: Threat Hunter original report, page 19. Redrawn in English from the original figure.

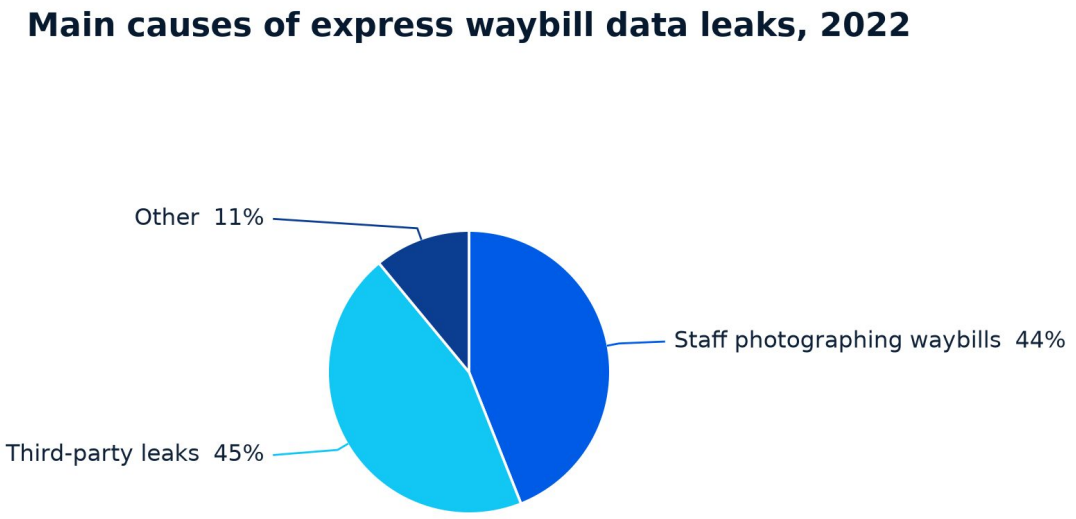
1. Artificial filming and leakage: courier site staff filmed the receipts and sold them to threat actors;

2. Leakage from third-party channels: The third-party channels used by enterprises store a large amount of express delivery data, and relevant third-party platforms have

Improper data storage or attacks by threat actors may lead to data leakage.

A certain cloud warehouse platform in the logistics industry was implanted with Trojans and exploited by vulnerabilities:

In 2022, the Threat Hunter intelligence platform monitored multiple organized groups of threat actors, targeting computers that print orders in cloud warehouses or e-commerce parks to carry out crimes. The threat actors gang falsely joined a domestic cloud warehouse on the pretext of applying for a job, and took the opportunity to install Trojan software on their work computers or printer computers, steal users' sensitive information and data, and sell them on the data trading market at a unit price of 3 yuan.



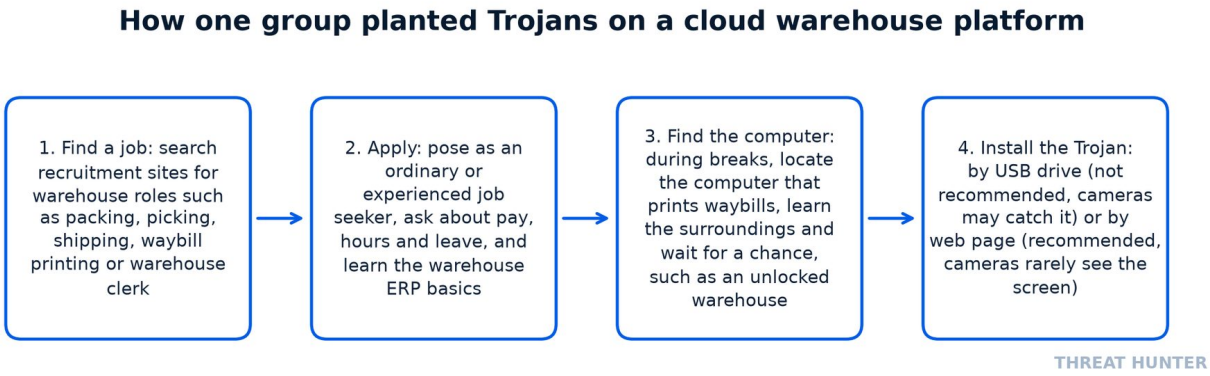
Main causes of express waybill data leaks in 2022: third-party leaks 45%, staff photographing waybills 44%

Source: Threat Hunter original report, page 20. Redrawn in English from the original figure.

3. A total of nearly 500 data leakage incidents in the e-commerce industry have been captured, involving well-known domestic and foreign brands.

In recent years, the development of new e-commerce and new retail has been in full swing. Many users shop through merchant-owned e-commerce, third-party e-commerce, live-streaming e-commerce, etc., resulting in a massive amount of shopping information, and the subsequent risk of data leakage has gradually increased. In 2022, Threat Hunter captured nearly 500 data breaches related to the e-commerce industry, involving well-known domestic and foreign brands such as beauty and clothing.

In May 2022, Threat Hunter detected that the e-commerce platform shopping data of a large domestic beauty brand was leaked, with more than 150,000 items leaked. The leaked data included customer names, phone numbers, product information, order time, payment methods, delivery addresses and many other sensitive information.



How one group planted Trojans on a cloud warehouse platform

Source: Threat Hunter original report, page 20. Redrawn in English from the original figure.

After analysis, the main reasons for data leakage in the e-commerce industry include:

1. Leakage in logistics links;

2. The "insider" in store operation was leaked;

3. Leakage of third-party tool software.

As mentioned earlier, there are many incidents of data leakage caused by third-party channels in the financial and logistics industries. In the e-commerce industry, many store merchants often use third-party development tools and software to manage customers, orders, services, etc. in order to improve operational efficiency. These tools have gradually become an important breakthrough for threat actors to steal data.

Take the third-party tool used by a certain store merchant as an example:

In March 2022, the Threat Hunter intelligence platform captured a third-party tool used by store merchants. The tool has a built-in back-end interface that can decrypt encrypted shopping data to obtain clear text order numbers, customer names, and addresses.

address, phone number and other information. Once this interface is maliciously attacked by threat actors, all merchants using this tool will have their order information leaked:

A store merchant caused information crawling due to third-party tools:

In June 2022, Threat Hunter used honeypot technology to capture an attack in which threat actors stole merchant tokens and crawled order information.



A third-party merchant tool whose back end exposes customer order details

Source: Threat Hunter original report, page 22. Original image retained; numbered translations added by Threat Hunter.

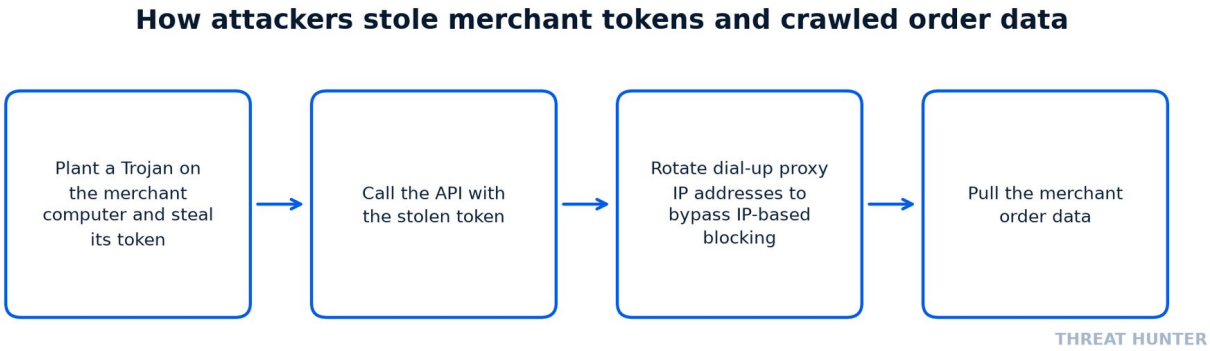
Figure guide

1	Authorized stores
2	Export table
3	Store name
4	Order number
5	Product name
6	Amount payable
7	Address
8	Expiry: 2022-03-12
9	Product: handwritten draft paper

Merchants need to authorize third-party tools to log in to the e-commerce platform, and cybercriminal groups successfully obtained the login tokens of multiple merchants by implanting backdoors in third-party tools, and used the obtained tokens to batch request order interfaces and crawl data. In order to avoid being discovered, threat actors use threat-actor resources such as instant dial proxy IP. Different businesses use different IPs.

4. A total of more than 200 data leakage incidents in traditional industries have been captured, involving 24 domestic and foreign well-known companies.

In 2022, Threat Hunter monitored more than 200 data leaks related to traditional industries. It is worth noting that data leaks in the automobile industry involved a total of 24 well-known domestic and foreign car brands. The leaked data included vehicle brands, identification codes, engine numbers, customer names, phone numbers, license plate numbers and other sensitive information. Among them, the highest leakage magnitude exceeded 2.7 million, and threat actors traded on the underground black market at an average unit price of 6.5 yuan.



How attackers stole merchant tokens and crawled order data

Source: Threat Hunter original report, page 22. Redrawn in English from the original figure.

Due to the late start of digitalization in traditional industries, there are often defensive weak links in data security construction. Among them, security flaws in API interfaces are the most typical weak links. Hackers use API security flaws such as unauthorized access and excessive exposure of sensitive data to launch data traversal attacks on API interfaces and steal user data or business data from the platform in batches.

Take the API attack incident of a large domestic enterprise as an example:

In December 2022, the Threat Hunter intelligence platform captured an extremely serious data breach in traditional industries. The API interface of a large domestic enterprise has a flaw in over-exposure of sensitive data. threat actors obtained sensitive information such as the names, ID cards, phone numbers, addresses, and passwords of employees within the enterprise by attacking the interface (as shown below).

The password obtained by threat actors is not the plaintext password, but the password md5. However, as long as the password complexity is not high, the original plaintext password can basically be restored through the rainbow table. Attackers can use the leaked phone number and password to log in to the company's OA system as an employee, invade the internal network and carry out further attacks, causing unpredictable losses to the company.

Note: If you want to know more data leakage cases related to API security, you can click to view the "2022 API Security Research Report".

Data protection measures and recommendations

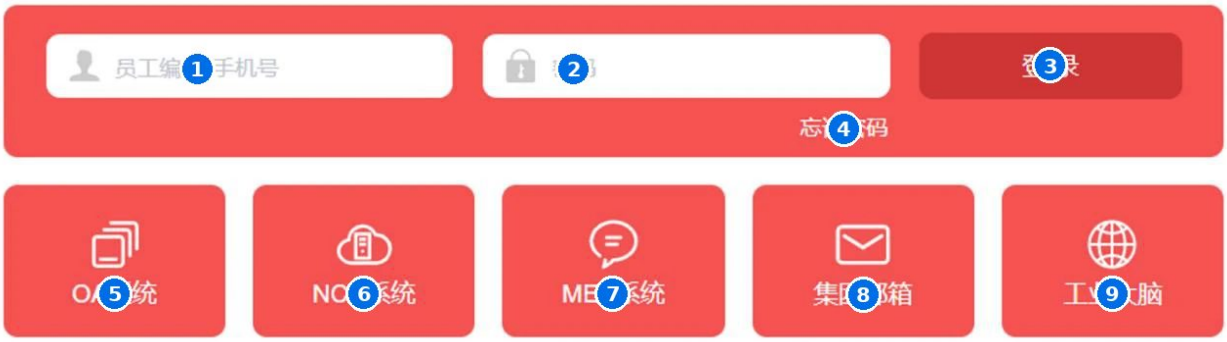


Employee data returned by the vulnerable API, including the password hash

Source: Threat Hunter original report, page 24. Original image retained; numbered translations added by Threat Hunter.

Figure guide

1	Address: Lanxi City ...
2	ID type: ID card
3	Employee name (masked)
4	Organization: ... Co., Ltd.
5	Department: Supply ... Division 1
6	Position: business supervisor



The affected enterprise's unified login portal for its internal systems

Source: Threat Hunter original report, page 24. Original image retained; numbered translations added by Threat Hunter.

Figure guide

1	Employee ID / phone number
2	Password
3	Log in
4	Forgot password
5	OA system
6	Finance system
7	Manufacturing system
8	Group email
9	Industrial analytics platform

4. Data protection measures and suggestions

In the increasingly severe security situation, data security has received increasing attention. In 2022, a series of data security policies and regulations such as the "Data Outbound Security Assessment Methods", "Data Security Management and Protection Guidelines", and "Information Security Technology Critical Information Infrastructure Security Protection Requirements" will be released one after another, proposing clearer requirements and implementation paths in terms of data security and compliance, which also increases the pressure on corporate security compliance.

In July 2022, the "Data Cross-border Security Assessment Measures" were reviewed and approved, proposing that data cross-border security assessment should adhere to the principles of combining prior assessment with continuous supervision, and combining risk self-assessment with security assessment.

In December 2022, the "Guidelines for Data Security Management and Protection of the Securities and Futures Industry" were released, providing guidance from the basic principles of data security management, organizational structure, systems, technology and other aspects, standardizing the data security management and protection work of industry organizations, reflecting the importance and urgency of data security for the securities and futures industry.

In October 2022, the "Information Security Technology Critical Information Infrastructure Security Protection Requirements" was released, stipulating the security requirements for critical information infrastructure operators in terms of identification and analysis, security protection, detection and evaluation, monitoring and early warning, active defense, and incident handling.

"Cyber attack" is one foot higher, and "data protection" needs to be even higher. Facing more and more data leakage risks, enterprises need to ensure the data security of the enterprise and its users from multiple dimensions. Not only must they have a clear understanding of the status of internal data assets, but they also need to detect and defend early and in a timely manner based on external risk intelligence monitoring.

For internal asset management, Threat Hunter security experts recommend that enterprises, on the basis of "business priority", conduct an overall combing of online APIs and flowing data assets on the API based on intelligence to achieve visibility of all API assets and flowing sensitive data, and then conduct continuous API defect assessment and attack threat awareness to achieve controllable API risks.

For external intelligence monitoring, enterprises can promptly perceive the risk of data leakage through multi-channel monitoring of the entire network, prevent and block incidents before they occur, discover the attack surface before attackers, and use monitoring of external intelligence to better protect the organization's digital assets.

Only by taking into account the internal and external perspectives of data security and combining internal data security with external threat intelligence monitoring can enterprises move steadily and quickly on the road to digital construction and innovative development.