SCout: Prying Into Supply Chains via a Public Query Interface

Liangxing Liu, Weili Han, *Member, IEEE*, Tao Zhou, and Xinyi Zhang

***Abstract*—The distribution network, including its flow information, in a supply chain system is usually a business secret to ensure the supply chain security and hold on to a favorable position in commercial competition. When more and more organizations deploy tracking systems to facilitate users, most of them focus much on the business growth but ignore the protection for the secrets. This paper therefore shows how we can pry into supply chains based on publicly acquired data via a public query interface. We design SCout, which crawls messages in social network services (SNSs) to acquire tracking numbers of an express company, and automatically retrieve the supply information from a public query interface, and then set up the distribution network of the target express company. SCout can also provide the flow information between any two distribution points. Furthermore, based on these obtained data, we analyze the relationship between the number of tracking numbers and the information of a distribution network. These experiments show that some express companies need to improve their awareness of data security. In particular, poor coding rules of tracking numbers can help adversaries obtain more tracking numbers easily. Thus, we provide some security countermeasures for express companies to defend from the above snooping. To the best of our knowledge, this paper is the first research to study the data security issue of logistics query systems from the business aspect.**

***Index Terms*—Internet of things, query interface, snooping, supply chain.**

I. Introduction

IN a globally challenging and competitive environment, the ability of supply chain management is critical for companies to become or remain competitive in business [1], [2]. With the help of the technologies of the Internet of things [3], [4], such as Radio-Frequency IDentification (RFID) and its supporting systems, the productivity of companies can be increased by making the supply chains run faster. These companies, in particular, provide public query interfaces to facilitate users in tracking merchandises to be delivered.

Security is an important issue in supply chain management and has gained more and more attention in recent years [5]–[9]. Most of these works only focus on traditional supply chain security concerns, particularly physical security, e.g., controlling theft and reducing contraband such as illegal drugs, illegal immigrants, and export of stolen goods [6], rather than data security. Moreover, after September 11, 2001, several researches are aware that the threat of terrorist attacks could be also related to supply chain security [6]–[8].

In this paper, we investigate how serious the problem of data leakage is when a company deploys a tracking system with a public query interface. Based on our work, the supply chain structure and flow of the company can be rebuilt when an adversary uses the information from social media and the interface. The supply chain structure of a company, including the distribution network and the number and location of warehouses and distribution centers, is usually unknown to end users and competing companies. It is usually a business secret to ensure supply chain security and to keep competitive in business. Therefore, the leakage of such important information may lead to great economic losses when a competing company can copy its gold lines.

We use express companies as examples, which have developed very fast in recent decades. Each express company usually offers a unique identifier, called tracking number, for every delivered package. We observe that some people write down their tracking numbers on social network services (SNSs) when they have delivered something. We can thereby acquire these tracking numbers. Furthermore, we find that express companies usually encode tracking numbers according to some preset rules. We try to get more tracking numbers automatically by a programmable tracking number generator. Based on the above data sources, we can pry into supply chains and know how the supply chains work.

The main contributions of this paper are as follows.

1) We articulate a new data security issue in supply chain management. We focus on a data security issue where an adversary can pry into supply chains via a public query interface. Much sensitive information, e.g., supply chain structure and chain's flow, can be leaked to an adversary, e.g., a business competitor.
2) We design and implement a tool called SCout, which can automatically reconstruct the supply chain network according to publicly acquired data. SCout acquires tracking numbers on SNSs and further generates more tracking numbers based on those acquired tracking numbers. Then, SCout analyzes the response data returned by a public query interface to pry into the supply chain. As a result, SCout sheds light on the data security issue in supply chain management.

Manuscript received November 29, 2013; revised March 5, 2014 and April 25, 2014; accepted June 21, 2014. This work was supported in part by the 12th Five-Year National Development Foundation for Cryptography under Grant MMJJ201301008, by CNNIC DNSLab, by the Natural Science Foundation of Shanghai under Grant 12ZR1402600, and by the Innovation Foundation of STCSM under Grant 12511504200.

The authors are with the Software School, Fudan University, Shanghai 201203, China, and also with Shanghai Key Laboratory of Data Science, Fudan University, Shanghai 201203, China (e-mail: 12212010009@fudan.edu.cn; wlhan@fudan.edu.cn; 12212010042@fudan.edu.cn; 10302010070@fudan.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSYST.2014.2337519

TABLE I
TRACKING NUMBER FORMATS FOR EXPRESS COMPANIES

Company	Format	Meanings	Example
S.F.Express (China)	xxx yyyyyyyy c	xxx: shipping address yyyyyyyy: serial number c: check digit	028 37622086 3
EMS (China)	E x yyyyyyyyy zz	E: EMS x: A∼Z, group yyyyyyyyy: serial number zz: CN or CS, country code	E Y 679411871 CS
UPS (United States)	1Z xxxxxx yy zzzzzzz c	1Z: UPS xxxxxx: alphanumeric account number of shipper yy: UPS service level code (e.g., 03 for United States Ground Express) zzzzzzz: package identifier c: check digit	1Z 311950 04 7808695 5
FedEx (United States)	xxxxxxxxxxxx	12 digits	802418695520
DHL (United States)	xxxxxxxxxx	10 digits	7266974046

The rest of this paper is organized as follows. Section II introduces some preliminary knowledge. Section III describes the design and implementation of SCout. We then present our experimental processes and evaluation results in Section IV. Next, we analyze experimental results and discuss security countermeasures for express companies in Section V. Section VI introduces the related work. Finally, Section VII concludes our work in this paper and introduces an outline of our future work.

II. BACKGROUND AND MOTIVATION

A. Supply Chain Management

In 1994, The International Center for Competitive Excellence defined supply chain management as follows: the integration of business processes from the end user through original suppliers that provides products services and information that add value for customers [10]. The concept of supply chain management has received increasing attention as modern business has entered the era that individual businesses no longer compete as solely autonomous entities but rather as supply chains [1]. Successful supply chain management can greatly save costs, improve efficiency, thus gain profits for companies.

For an express company, a supply chain consists of the following two basic components.

1) Point: A point, usually a distribution center, holds information such as name, location, flow.
2) Line: A line, a route between two distribution centers, holds information like the package flow between two distribution centers.

B. Investigation of Tracking Number Formats

Express companies offer a tracking number for each package in general. It works as an identifier for each package. Customers can track their packages via official tracking systems or a public query interface and know exactly when and where the packages go through.

Different companies adopt different tracking number formats when they set tracking numbers. We make a survey about how express companies encode their tracking numbers. We investigate several popular express companies in China and the USA, including S.F. Express, EMS, FedEx, UPS, and DHL. For S.F. Express, a tracking number consists of 12 digits. Generally, the first three digits indicate the shipping address of packages, and the middle eight digits are serial numbers. The last digit is a check digit. An example of S.F. Express tracking number is *028376220863*.

Table I shows common tracking number formats for different express companies and lists relevant examples. Based on our analysis, the number of S.F. Express is the easiest for an adversary to guess using a known tracking number as a seed.

C. Motivation

The supply chain network plays an important role in ensuring the reliability of supply chain service, which is the most important for a productive supply chain [11]. As a case, DHL is a market leader in the supply chain industry in Southeast Asia. In Singapore, on May 28, 2013, DHL announced its planned investment for its business in Southeast Asia to a tune of €140 million by 2015. The bulk of the investments will go toward investing in new facilities, advanced IT solutions, expansion in transportation, and bolstering staff strength and training. These investments will further enhance DHL supply chain's market-leading position across the key consumer, retail, automotive, and technology industries and grow its business in emerging sectors [12].

Because of its supply chain network, an express company holds on to a favorable position in competition with other express companies. Thus, the detailed supply chain network is very important for an express company for two reasons.

1) For a competing company, when it wants to enter a new region, it can copy the supply chain network of an existing company and know how to deploy distribution network.
2) For an internal point in the chain, if the manager can know the detailed flow of other similar points, he or she could bargain with the headquarter for his or her cost, particularly when the internal point is assigned an independent financial budget.

In addition, if the supply chain network is serving for a retailer, the data disclosure can cause the following hypothetical attack case:

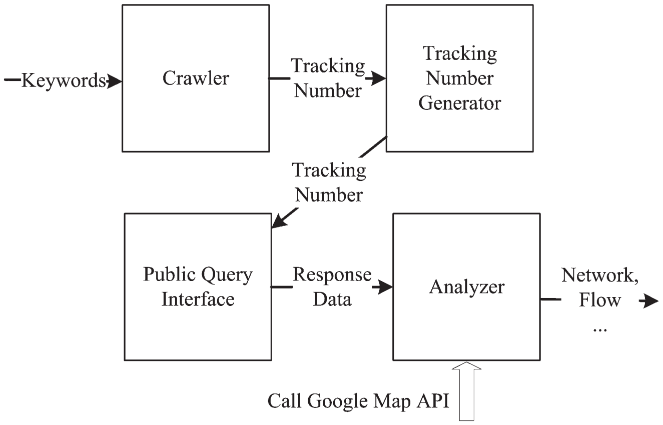


Fig. 1. Key flow of SCout.

> A retailer R is selling a high-profit merchandise, which can offer a tracking service to any end consumer. Its commercial adversaries want to know its statistic information of sales on the merchandise. Adversaries can assign its employees to act as consumers to buy the merchandise for a period. Based on a simple statistic method, they may know the sales information, including the portion of sales from different suppliers, which are usually the top commercial secrets of the retailer R.

This paper, therefore, is motivated by these issues and then empirically studies how an adversary can pry into supply chains based on publicly acquired data via a public query interface.

Note that we use S.F. Express as a case because it is a popular express service in China and the data can be publicly acquired more easily. The issues in S.F. Express also exist in other express companies and commercial supply systems.

III. Design of SCout

SCout is designed based on two facts: First, there are messages on SNSs that contain tracking numbers; second, there exist rules when coding tracking numbers. Therefore, it is possible to write a program to produce massive new tracking numbers using crawled tracking numbers.

A. Key Flow of SCout

The key flow of SCout is shown in Fig. 1. There are four basic modules: *Crawler*, *Tracking Number Generator*, *Public Query Interface*, and *Analyzer*. The *Crawler* is responsible for crawling messages containing input keywords on SNSs to acquire tracking numbers. *Tracking Number Generator* produces related tracking numbers according to coding rules. Through a *Public Query Interface*, we can get shipping information of packages. *Analyzer* further analyzes these response data to crack into supply chains.

B. Key Algorithms

1) Crawler: The crawler is a Sina Weibo crawler. Sina Weibo (http://weibo.com/) is a popular Chinese SNS used by well over 30% of Internet users in China. As a hybrid of Twitter and Facebook, about 100 million messages are posted each day on Sina Weibo [13], [14], recording daily lives of the users. Among the massive messages, there are some messages that contain tracking numbers of express companies.

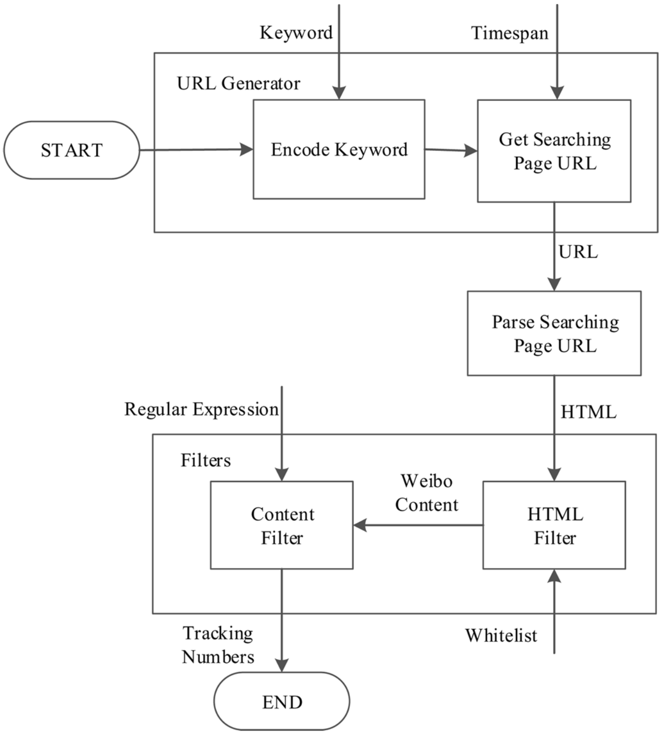


Fig. 2. Key flow of Sina Weibo Crawler.

The key flow of Sina Weibo crawler is shown in Fig. 2. To acquire tracking numbers, the crawler carries out the following four main steps: 1) generate a searching page URL by setting keywords and time span (restrict weibo messages to a specific time interval); 2) parse the searching page URL to get the HTML content; 3) filter the HTML content to get *weibo* content; and 4) get validated tracking numbers according to formats.

2) Tracking Number Generator: Tracking Number Generator takes advantage of original tracking numbers acquired by Sina Weibo Crawler to generate quantities of new tracking numbers, which are tracking numbers of S.F. Express in the later experiments in Section IV.

As aforementioned in Section II, the middle eight digits in an S.F. Express tracking number are serial numbers, and the last digit is a check digit. We collect many tracking numbers on the Internet and make a further comparison among these tracking numbers. We find that the ninth, tenth, and eleventh digits affect the check digit in different ways. For example, when generating a new tracking number with the same ninth digit as the original tracking number, if the tenth digit is 3 or 6, and the eleventh digit is 9, the last check digit is dealt with in the following way.

1) If the check digit in the original tracking number is over 5, the check digit minus 5 is the new check digit in the new tracking number.
2) If the check digit in the original tracking number is less than 5, the check digit plus 5 is the new check digit in the new tracking number.

TABLE II
RULES FOR CREATING NEW TRACKING NUMBERS

No.	Conditions				Value of N_{12}
1	$N_9 - n_9 = 1$	$n_9\%2 = 1$		$n_{12} - 8 > 0$	$n_{12} - 8$
2				$n_{12} - 8 < 0$	$n_{12} + 2$
3		$n_9\%2 = 0$		$n_{12} - 7 > 0$	$n_{12} - 7$
4				$n_{12} - 7 < 0$	$n_{12} + 3$
5	$N_9 - n_9 \neq 1$	$n_{11} = 9$	$N_{10} = 3$ or $N_{10} = 6$	$n_{12} - 5 > 0$	$n_{12} - 5$
6				$n_{12} - 5 < 0$	$n_{12} + 5$
7			$N_{10} \neq 3$ and $N_{10} \neq 6$	$n_{12} - 4 > 0$	$n_{12} - 4$
8				$n_{12} - 4 < 0$	$n_{12} + 6$
9		$n_{11} \neq 9$		$n_{12} - 1 > 0$	$n_{12} - 1$
10				$n_{12} - 1 < 0$	$n_{12} + 9$

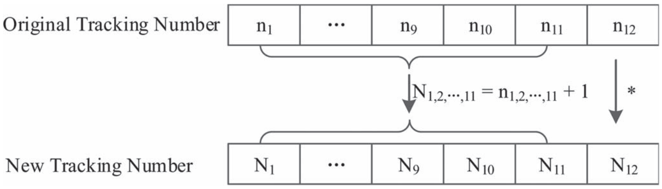


Fig. 3. Tracking number. (*) Table II shows the rules of generating new tracking numbers.

Table II shows the rules we find out through investigation for generating new tracking numbers of S.F. Express. Notations in Table II are shown in Fig. 3.

3) Public Query Interface: The *Public Query Interface* we use is a query interface provided by http://www.kuaidi100.com/. We use *httpclient* to acquire shipping information via the interface. The query interface returns responses in *JavaScript Object Notation (JSON)*, which is an open standard format to transmit data objects consisting of attribute–value pairs using human-readable text. A *JSON* object is an item of response written in *JSON*. The following is a *JSON* object.

```
{
  “message”:“ok”,
  “nu”:“025632311271”,
  “ischeck”:“1”,
  “com”:“shunfeng”,
  “updatetime”:“2013-05-20 11:20:52”,
  “status”:“200”,
  “condition”:“F00”,
  “data”:
   [
    {“time”:“2013-05-02 17:59:47”,
     “context”:“Shanghai Received”,
     “ftime”:“2013-05-02 17:59:47”}
    {“time”:“2013-05-03 10:21:34”,
     “context”:“Shanghai Delivering”,
     “ftime”:“2013-05-03 10:21:34”}
    {“time”:“2013-05-03 14:08:29”,
     “context”:“Shanghai Signed”,
     “ftime”:“2013-05-03 14:08:29”}
   ]
}
```

The meanings of the parameters in the *JSON* object are shown in Table III.

TABLE III
MEANINGS OF THE PARAMETERS IN THE JSON OBJECT

Parameter	Meaning
message	A text message returned by the query interface
nu	A tracking number
ischeck	Whether the package is received and signed
com	Express company
updatetime	Query time
status	Query status
condition	Package status at the moment
data	Detailed shipping information

4) Analyzer: The *Analyzer* needs the following data to construct a supply chain network:

1) the routes from the shipping address to the receiving address of packages;
2) the name, number, and location of distribution centers in each route.

Before constructing a supply chain network, the *Analyzer* needs to parse *data* field in the *JSON* object, which contains detailed shipping information, and then conclude the following data:

1) the routes from a distribution center to another one;
2) the flow of distribution centers;
3) the flow of routes between two distribution centers.

The *Analyzer* calls the *Google Map API* [15] to display distribution centers and draw routes on the map in the following way.

1) Every distribution center is translated to a location that consists of longitude and latitude. For example, [31.2303930, 121.4737040] is for *Shanghai*.
2) A dot is drawn on the Google Map corresponding to the location of a distribution center. Once a route of a tracking number from a shipping address to a receiving address goes through the distribution center, the flow of it adds one.
3) A line is drawn when the route from a distribution center to another one exists. Once a route of a tracking number from a shipping address to a receiving address includes this line, the flow of this line adds one.

The *Analyzer* traverses all data to construct a supply chain network on the Google Map. In addition, the flow information is shown on the corresponding dots and lines.

TABLE IV
PACKAGE STATUS AMONG THE 75 201 TRACKING NUMBERS OF PACKAGES

Condition	Meaning	Number
00	still in transportation	344
B00	without shipping information	31
C60	returned package	3
F00	ok	74,785
H100	still in distribution	35
F10	refuse to sign	3
Total		75,201

IV. EVALUATION

A. Crawling Tracking Numbers via SNSs

This experiment is to analyze the possibility of collecting tracking numbers on the Internet automatically. As aforementioned, the SNS we select is Sina Weibo, and in this experiment, we try to produce tracking numbers for S.F. Express. We conduct this experiment in the following way.

1) We set the searching *keywords* in *Sina Weibo Crawler* as *shunfeng+tracking number* (in simplified Chinese).
2) We set a *time interval* for *Sina Weibo Crawler*: from January 1, 2013 to July 31, 2013.

The crawler successfully gets 499 valid tracking numbers in our experiment.

> **Insight**: It is possible to collect tracking numbers of a specific express company on SNS. In addition, although the targeted express company is S.F. Express company, the crawler can be also employed to crawl tracking numbers of other express companies.

B. Analyzing the Effectiveness of Generating Tracking Numbers

With 499 original tracking numbers, the *Tracking Number Generator* produces a total of 268 517 distinct tracking numbers. SCout makes queries about these tracking numbers via the *Public Query Interface* and collects the response data. Among the 268 517 tracking numbers, 75 201 items of response data are returned by the query interface. According to the value of parameter *condition* in response data, which indicates the package status, we divide them into different categories.

As we can see, about 28% of tracking numbers are returned with valid shipping information by the query interface, whereas others encounter query errors. Not all tracking numbers generated by *Tracking Number Generator* are valid for the following reasons.

1) The query interface usually restricts time interval of queries of tracking numbers. Information about tracking numbers for packages, which were delivered three months ago, cannot be acquired via the query interface.
2) Newly generated tracking numbers have not been assigned to packages yet. In other words, these tracking numbers, which could be real, are not in use at the moment.

In Table IV, 74 785 items of packages have completed shipping information. However, not all these response data can be used for our experiment. For example, some items of response data only contain the name of the person who signs the package, which we cannot use to get the shipping address, receiving address, and transportation route. As a result, we further add up the number of response data that contains information, e.g., shipping address and receiving address. The total number is 54 126, which means that these are response data we can use to conduct our experiments to pry into supply chains.

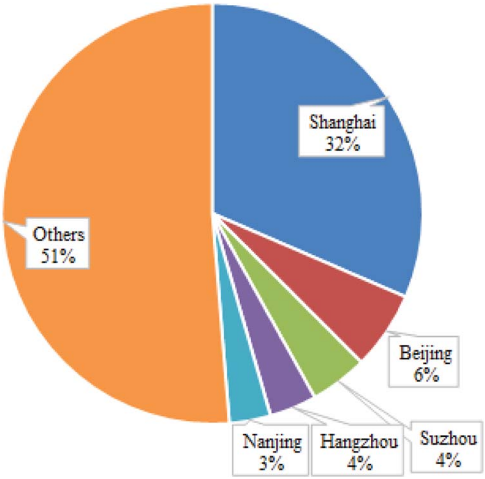


Fig. 4. Distribution of receiving addresses for packages shipped from *Shanghai*. About 32% packages are delivered to *Shanghai* locally.

> **Insight**: From the above statistics, we find that with 499 original tracking numbers, it is possible to generate 268 517 tracking numbers in total, among which 54 126 items of response data are usable finally. The percentage is about 20%. As a result, SCout proves to be effective in producing new tracking numbers.

C. Prying Into Supply Chains

1) Flow of the Route From a Shipping Address to a Receiving Address: For every shipping address, we can find out the number of packages to a certain receiving address. Take *Shanghai* as an example, we count the number of packages to each receiving address, and Fig. 4 shows the percentage of receiving addresses for packages shipped from *Shanghai*.

> **Insight**: According to the above statistics, we find out that the number of packages delivered to the same receiving address as the shipping address is a lot more than that of other addresses.

2) Flow of Every Distribution Center: Response data for every tracking number includes shipping information, consisting of every distribution center. Once a package goes through the distribution center, we add one to the value of the flow of this distribution center.

In total, we find out 346 distribution centers in our experiment and count the flow of every distribution center, i.e., the number of packages that go through this distribution center. We divide the distribution center into five categories according to the percentage of packages that go through a distribution center against the total number of packages. Results are shown in Table V.

The top four distribution centers in scale are: Shenzhen, Beijing, Guangzhou, and Shanghai.

TABLE V
FIVE DISTRIBUTION CENTER CATEGORIES

Category	Percentage	Number
Very Large	>10%	4
Large	2-10%	30
Midum	0.2-2%	78
Small	0.02-0.2%	159
Very Small	≤0.02%	75

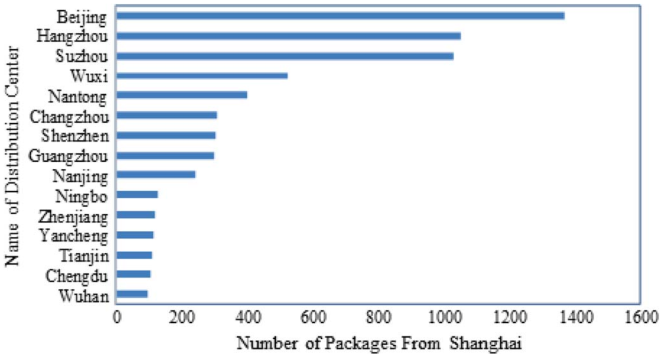


Fig. 5. Number of packages from *Shanghai* to other distribution centers. Most of the top 15 distribution centers with the largest flow are in *Jiangsu* province and *Zhejiang* province.

3) Flow of the Route Between Two Distribution Centers: Once a package goes from one distribution center to another distribution center, we add one to the value of the flow of the route between two distribution centers. The route between two distribution centers differentiates from the route from a shipping address to a receiving address in the following way: The former indicates a real line of two adjacent points during the transportation process, whereas the latter indicates a virtual line connecting a pair of start point and end point.

Take *Shanghai* as an example again, the top 15 distribution centers with the largest flow from *Shanghai* are shown in Fig. 5.

As shown in Fig. 5, among the top 15 distribution centers with the largest flow from *Shanghai*, 9 distribution centers are in *Jiangsu* province and *Zhejiang* province, which is not a surprise because of the existence of the Economic Circle of the Yangtze River Delta.

4) Network of the Supply Chain: For each package, there is a route from its shipping address to the receiving address. In our experiment, we find out a total of 7789 routes. Before a package arrives at the receiving address finally, it may go through several intermediate nodes called transfer station.

First, we count how many intermediate nodes a route includes. Here, if the value is -1, the shipping and receiving addresses are the same; if the value is *0*, the package is directly sent to the receiving address from the shipping address without going through any transfer stations. Fig. 6 shows the number of routes with corresponding number of intermediate nodes.

As shown in Fig. 6, there are one or two intermediate nodes before a package arrives at the receiving address in most cases, with a percentage close to 70%.

Second, we count the number of possible transportation lines between two addresses. For example, from *Shanghai* to *Beijing*, we find out that there are 5 different transportation lines: *Shanghai* to *Guangzhou* directly; *Shanghai* to *Guangzhou* via *Shenzhen*; *Shanghai* to *Beijing* via *Hangzhou*; *Shanghai* to *Beijing* via *Hangzhou* and *Shenzhen*; and *Shanghai* to *Guangzhou* via *Wuxi*, *Hangzhou*, and *Shenzhen*. By comparing the flow of each possible transportation line, we can know which line is most often used. We further count how many routes are possible for a package with the same shipping and receiving address. Fig. 7 shows the total number of times with the corresponding number of routes from a shipping address to a receiving address.

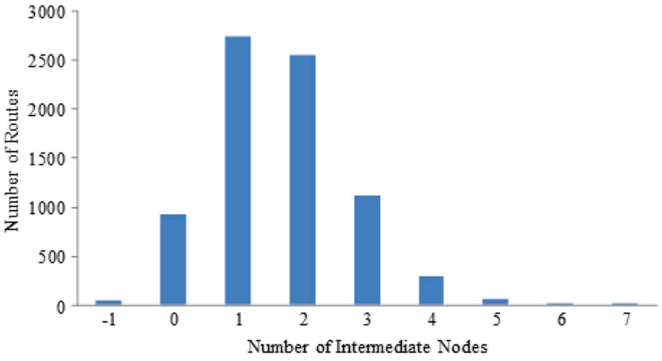


Fig. 6. Number of routes with corresponding number of intermediate nodes. Each bar represents the number of routes that include the corresponding number of intermediate nodes. For most routes among these 7789 routes, there are 1–2 intermediate nodes. Note that there are at most 7 intermediate nodes here.

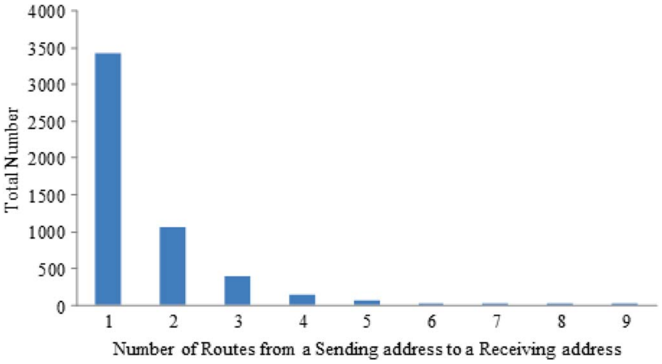


Fig. 7. Total number of times with the corresponding number of routes from a shipping address to a receiving address. Each bar represents the total number of times when there are a corresponding number of transportation lines from a shipping address to a receiving address. For most packages with the same shipping and receiving address, there is a unique route. There are at most 9 routes from a shipping address to a receiving address found out in our experiment.

As shown in Fig. 7, for most packages with the same shipping and receiving address, there is only one possible transportation line. In our experiment, we find out that there are at most 9 routes from a shipping address to a receiving address.

Third, by taking advantage of information we get using SCout, we can build up a network of the supply chain.

Fig. 8 shows a part of a supply chain network. It includes 5 possible transportation lines for packages from *Shanghai* to *Guangzhou*, and each transportation line is marked with its flow, respectively.

V. DISCUSSION

A. Correctness of SCout

The original tracking numbers are crawled randomly in Sina Weibo. Moreover, new tracking numbers are produced with the

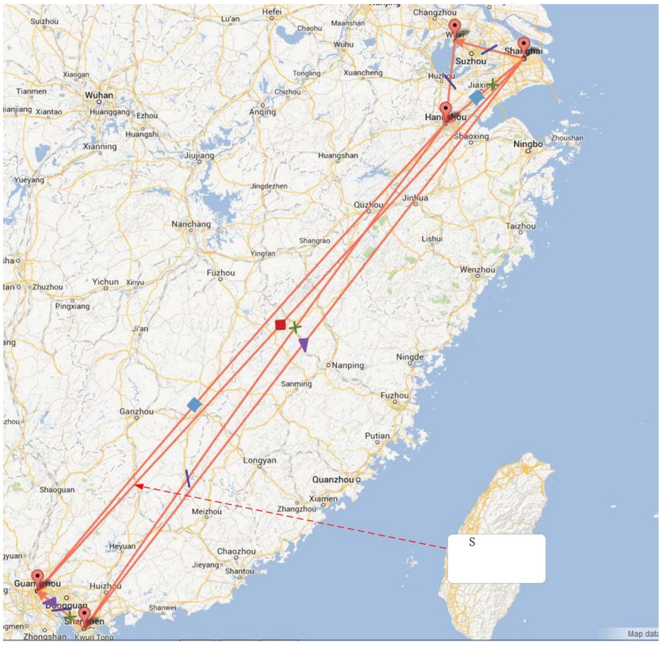


Fig. 8. Part of the supply chain network. It shows five different transportation lines and their flows for packages from *Shanghai* to *Guangzhou*. Lines with the same marks belong to the same route.

same shipping address as the original number. Therefore, the composition of original tracking numbers has a great impact on the experimental results.

The experimental results on the flow information do not necessarily imply the true flow of every distribution center and route. This issue is led by the following two reasons.

1) As the difference in popularity of Sina Weibo among regions all over China, it is less probable to get tracking numbers sent from small cities. Therefore, as shown in the experiments in Section IV, by SCout, while we can find out a general flow of distribution centers and routes, some distortion exists.
2) The results indicate relative flow of distribution centers and routes between two points. Although we try to generate new tracking numbers as many as possible with an original tracking number, it is impossible to generate tracking numbers exhaustively, as we only find out partial rules for generating tracking numbers. As a result, data in our experiments reflect a relative amount of flow.

However, when enough original tracking numbers are crawled, the results would be closer to actual situations.

B. Impact of the Number of Tracking Numbers on Experimental Results

We conduct an experiment to find out the impact of the number of tracking numbers we use in experiments on the structure of a supply chain reconstructed by SCout. There are three metrics to measure the completeness of tracking a supply chain, namely, number of routes from a shipping address to a receiving address; number of routes between two distribution centers; and number of distribution centers.

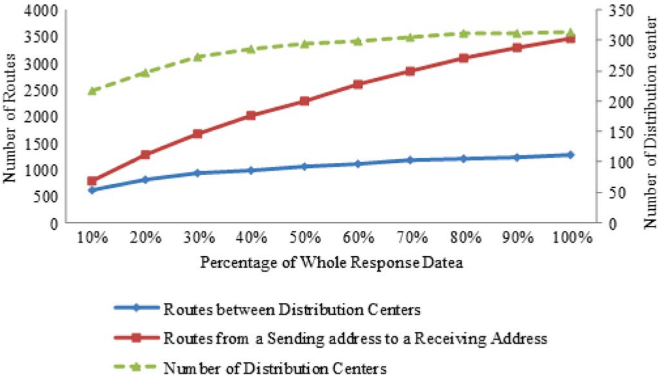


Fig. 9. Comparison of the number of routes and distribution centers that SCout finds out with a percentage of complete response data. The increase rate of both curves tends to decline when the percentages grow.

We randomly take a total of 20 000 [1] from the valid 54 126 tracking numbers as a sample set and further randomly select a certain percentage of the whole response data. We conduct several experiments to count the number of routes and distribution centers SCout finds out with different numbers of tracking numbers. Fig. 9 shows the number of routes from a shipping address to a receiving address, routes between distribution centers, and the number of distribution centers when the amount of response data varies. In Fig. 9, the left Y-axis represents the number of routes, and the right Y-axis represents the number of distribution centers.

As shown in Fig. 9, the increasing rate of both curves tends to decline when the percentages grow, and we use the method of regression analysis based on the natural logarithm to predict the impact of the number of tracking numbers on the supply chain network. The relationship between the number of distribution centers that SCout finds out, denoted as N_{center}, and the number of tracking numbers, denoted as N_{number}, is as follows:

$$N_{\text{center}} = 43.445 \ln\left(\frac{N_{\text{number}}}{2000}\right) + 220.08.$$

> **Insight**: Using this formula, we can predict the number of distribution centers SCout can find out with a specific number of tracking numbers.

C. Risk When Tracking Numbers Containing Shipper Identifiers

Several express companies such as UPS will encode a shipper identifier into their tracking numbers. That is dangerous for both shippers and customers. For example, when a customer shows his or her UPS's tracking number after he or she buys a special merchandise, his or her buying behaviors may be disclosed due to the combination based on the identifier. An adversary can know the tracking numbers of the other customers from their SNSs or automatically generate them from a seed.

[1] This step can save the data processing time. At the same time, the results can show the features of the whole of the valid 54 126 tracking number.

D. Security Countermeasures

Although our experiments are conducted within S.F. Express, the results have a common implication on similar logistics companies. These companies usually offer a public query interface for customers to make queries about delivery information related to their purchased products. Thus, we propose the following countermeasures to resolve the data security issue.

1) Securer coding for the tracking number. Insecure coding rules amplify the effectiveness of SCout. Furthermore, there is risk when tracking numbers containing shipper identifiers. Thus, it is very important for a tracking system to securely encode their track numbers. First, the tracking number cannot be guessable by, e.g., encrypting a key part of sensitive tracking numbers, even when an adversary has some seed tracking numbers. Second, the tracking number should not contain some identifier information, including shipper identifiers, and possibly receiver identifiers.
2) Controlling the data queries. Based on our experiment, we can collect many valid tracking numbers from the current social media. Then, an adversary can collect the supply chain information from the data queries. Thus, to mitigate risks, the information provider should control the data queries as follows.1) Cutting the time windows for query. Once the package is delivered, the data should be closed to the public.2) Blurring information that a public query interface returns. Some detailed transportation information about packages, such as concrete location, should be blurred to counter the risk mentioned in this paper, e.g., changing from *East Nanjing Road* to *Exxx Road*.3) Controlling batch queries. The server may take measures to ban the *IP address* if massive requests from the same *IP address* are found.
3) Strengthening the risk concern of common customers. Many customers like showing their tracking numbers in SNSs. According to the research in this paper, it should be dangerous for the logistics companies and other people. Thus, it is necessary to tell them not to show their tracking numbers to the public.
4) Forbidding the underground trading of tracking numbers. When we search “selling tracking numbers” in Google, we can see that many websites are selling tracking numbers. Through our investigation, we find out that it is possible to buy large quantities of tracking numbers on the Internet. In fact, the underground trading of tracking numbers has become a business chain. Staffs within express companies could sell tracking numbers for profits, and thus, information about customers is leaked.
5) Enhancing the security of supporting systems in supply chains. If an adversary can bypass the safeguards in the public query interfaces, even intrude the inner network, the adversary can obtain more information. This is a serious threat in the current cyberworld. As a result, it is necessary to enhance the security of supporting systems in supply chains.

While these countermeasures above can strengthen supply chain security, they may cause other problems. For example, more complicated and securer coding for tracking numbers will increase query time of transaction. Therefore, an express company has to make a tradeoff between coding complexity and time cost.

VI. Related Work

The term *Supply Chain Management* was originally introduced by consultants in the early 1980s [16]. Since 1990s, more and more attention has been paid to this field. Cooper *et al.* differentiated *Supply Chain Management* from *Logistics Management* in the literature [10] and proposed a conceptual model for the supply chain management. In the literature [17], in order to successfully implement supply chain management, Lambert *et al.* concentrated on operationalizing the supply chain management framework suggested in the literature [10].

Supply chain security is a very important issue in supply chain management. In the previous work, supply chain security mainly focused on physical threats. In the literature [5], Lee and Wolfe pointed out that supply chain managers today face a dilemma: How to improve security without jeopardizing supply chain effectiveness. They believed that it is possible to create strategies that both prevent and mitigate security breaches while still maintaining productivity. In the literature [6], Lee and Whang mentioned that, after September 11, 2001, the threat of terrorist attacks has heightened the need to assure supply chain security. They described how the principles of total quality management can be actually used to design and operate processes to assure supply chain security. In the literature [7], certain *robust* strategies were presented that possess two properties. First, these strategies can enable a supply chain to manage the inherent fluctuations efficiently regardless of the occurrence of major disruptions. Second, these strategies can make a supply chain become more resilient in the face of major disruptions. The literature [8] looks at the twin corporate challenges: 1) preparing to deal with the aftermath of terrorist attacks; and 2) operating under heightened security. In addition, they looked at how companies should be organized to meet those challenges efficiently and suggests a new public–private partnership. Sheffi stated that while their work is focused on the USA, it has worldwide implications.

Security in RFID-based supply chain systems has been paid much attention. The literature [18] presents a brief description of RFID systems. Weis *et al.* described privacy and security risks and how they apply to the unique setting of low-cost RFID devices. In the literature [19], an overview of the current solutions to RFID security and privacy is given, and Gao *et al.* proposed a new approach that exploits randomized read access control and thus prevents hostile tracking and man-in-the-middle attack. The literature [20] proposes protocol *TRACKER* to combat counterfeiting of pharmaceutics or luxury objects, which becomes a major threat to supply chains today. *TRACKER* is used for object genuineness verification in RFID-based supply chains. More precisely, TRACKER allows to securely identify which (legitimate) path an object/tag has taken through a supply chain. In the literature [21], Li and Ding stated that existing RFID solutions cannot be applied directly in the context of supply chain management because of a set of special RFID security requirements to be addressed for supply chain management. Therefore, they identified the unique set of security requirements in supply chains and proposed a practical design of RFID communication protocols that satisfy the security requirements.

Previous works on supply chain security mainly focus on traditional security issues, particularly physical security, e.g., controlling theft, and terrorist attacks. We instead pay attention to a new data security issue in supply chain management. Based on our research, we find out that it is possible to generate large quantities of new tracking numbers based on original tracking numbers. Moreover, we can crawl tracking numbers on SNSs. We develop a tool called SCout to pry into supply chains and reconstruct the supply chain network of express companies.

VII. Conclusion and Future Work

To the best of our knowledge, this paper is the first research to study the data security issue of logistics query systems from the business aspect. We first articulated a new data security issue in supply chain management that it is possible for others to pry into the supply chain via a public query interface. In addition, we have developed a tool called SCout, which can automatically reconstruct supply chain based on publicly acquired data. When we use express companies as cases, we conducted several experiments to evaluate SCout. The experimental results showed that it is possible for SCout to pry into the supply chain. Based on the experimental results, we proposed some countermeasures on how to strengthen the security of the supply chain.

In our future work, we will further improve the efficiency of SCout in crawling tracking numbers and generating new tracking numbers for more commercial companies and discuss our results with these companies. Moreover, we will analyze the data in Scout and further study the privacy issue of customers. Finally, we want to implement our proposed countermeasures and empirically study the efficiency of some countermeasures, including the securer coding, the enhanced control on the query interface, and users' security awareness.

Acknowledgment

The authors would like to thank anonymous reviewers for their comments. Weili Han is the corresponding author.

References

[1] D. M. Lambert and M. C. Cooper, "Issues in supply chain management," *Ind. Market. Manage.*, vol. 29, no. 1, pp. 65–83, Jan. 2000.

[2] M. C. Cooper and L. M. Ellram, "Characteristics of supply chain management and the implications for purchasing and logistics strategy," *Int. J. Logist. Manage.*, vol. 4, no. 2, pp. 13–24, 1993.

[3] L. Atzoria, A. Ierab, and G. Morabitoc, "The Internet of things: A survey," *Comput. Netw.*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.

[4] L. Zheng *et al.*, "Technologies, applications, governance in the Internet of things," in *Internet of Things Global Technological and Societal Trends*. Aalborg, Denmark: River Publishers, 2011, pp. 141–176.

[5] H. L. Lee and M. Wolfe, "Supply chain security without tears," *Supply Chain Manage. Rev.*, vol. 7, no. 3, pp. 12–20, Jan./Feb. 2003.

[6] H. L. Lee and S. Whang, "Higher supply chain security with lower cost: Lessons from total quality management," *Int. J. Prod. Econ.*, vol. 96, no. 3, pp. 289–300, Jun. 2005.

[7] C. S. Tang, "Robust strategies for mitigating supply chain disruptions," *Int. J. Logist. Res. Appl.*, vol. 9, no. 1, pp. 33–45, Mar. 2006.

[8] Y. Sheffi, "Supply chain management under the threat of international terrorism," *Int. J. Logist. Manage.*, vol. 12, no. 2, pp. 1–11, 2001.

[9] F. Giannotti, L. Lakshmanan, A. Monreale, D. Pedreschi, and H. Wang, "Privacy-preserving mining of association rules from outsourced transaction databases," *IEEE Syst. J.*, vol. 7, no. 3, pp. 385–395, Sep. 2013.

[10] M. C. Cooper, D. M. Lambert, and J. D. Pagh, "Supply chain management: More than a new name for logistics," *Int. J. Logist. Manage.*, vol. 8, no. 1, pp. 1–14, 1997.

[11] B. Li, *A Study Of Critical Factors of Customer Satisfaction in Parcel Delivery Service*. Lincoln, NE, USA: Univ. of Nebraska Press, 2002.

[12] DHL to Invest Euro 140 Million to Enhance its Supply Chain Business In Southeast Asia, May 2013. [Online]. Available: http://en.prnasia.com/story/80237-0.shtml

[13] K. Rapoza, *China's Weibo vs US's twitter: And the Winner is?* May 2011. [Online]. Available: http://www.forbes.com/sites/kenrapoza/2011/05/17/chinas-weibos-vs-uss-twitter-and-the-winner-is.html

[14] L. Yu, S. Asur, and B. A. Huberman, "What trends in Chinese social media," *arXiv preprint arXiv:1107.3522*, 2011.

[15] M. N. Boulos, "Web GIS in practice iii: creating a simple interactive map of England's strategic health authorities using Google Maps API, Google Earth KML, MSN Virtual Earth map control," *Int. J. Health Geograph.*, vol. 4, no. 1, pp. 22–40, 2005.

[16] R. K. Oliver and M. D. Webber, "Supply-chain management: Logistics catches up with strategy," *Outlook*, vol. 5, no. 1, pp. 42–47, 1982.

[17] D. M. Lambert, M. C. Cooper, and J. D. Pagh, "Supply chain management: Implementation issues and research opportunities," *Int. J. Logist. Manage.*, vol. 9, no. 2, pp. 1–20, 1998.

[18] S. A. Weis, S. E. Sarma, R. L. Rivest, and D. W. Engels, "Security and privacy aspects of low-cost radio frequency identification systems," in *Security in Pervasive Computing*. Berlin, Germany: Springer-Verlag, 2004, pp. 201–212.

[19] X. Gao *et al.*, "An approach to security and privacy of RFID system for supply chain," in *Proc. IEEE Int. Conf. E-Commerce Technol. Dyn. E-Business*, 2004, pp. 164–168.

[20] E.-O. Blass, K. Elkhiyaoui, and R. Molva, "Tracker: Security and privacy for RFID-based supply chains," in *Proc. Netw. Distrib. Syst. Security Symp.*, San Diego, CA, USA, 2011.

[21] Y. Li and X. Ding, "Protecting RFID communications in supply chains," in *Proc. 2nd ACM Symp. Inf., Comput. Commun. Security*, 2007, pp. 234–241.

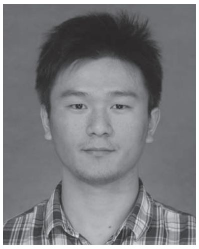
Liangxing Liu received the Bachelor's degree from East China Normal University, Shanghai, China, in 2012. He is currently working toward the Master's degree in the Software School, Fudan University, Shanghai.

His research interests mainly include policy-based management and information security.

Weili Han (M'08) received the Ph.D. degree in computer science and technology from Zhejiang University, Hangzhou, China, in 2003.

Then, he joined the faculty of the Software School, Fudan University, Shanghai, China, where he is currently an Associate Professor. From 2008 to 2009, he visited Purdue University, West Lafayette, IN, USA, as a Visiting Professor funded by China Scholarship Council and Purdue University. His research interests are mainly in the fields of policy-based management, Internet-of-things security, information security, and distributed systems.

Dr. Han is a member of the Association for Computing Machinery; the Special Interest Group on Security, Audit and Control; and the Chinese Computing Federation. He serves in several leading conferences and journals as Program Committee Members, Reviewers, and an Associate Editor.



Tao Zhou is currently working toward the Master's degree in the Software School, Fudan University, Shanghai, China.

He is working hard on the research topic on policy-based management and Android security.

Xinyi Zhang received the Bachelor's degree in the Software School, Fudan University, Shanghai, China, in 2014. She is going to the Systems, Algorithms, Networking and Data Laboratory, University of California, Santa Barbara, CA, USA, to pursue a Ph.D. degree.

Her research interests mainly include information security and data mining.