I've Got Your Packages: Harvesting Customers' Delivery Order Information using Package Tracking Number Enumeration Attacks

Simon S. Woo*, Hanbin Jang*, Woojoong Ji, Hyoungshick Kim†
Computer Science and Engineering Department, Sungkyunkwan University
{swoo,hanbin,woojoong,hyoung}@skku.edu

ABSTRACT

A package tracking number (PTN) is widely used to monitor and track a shipment. Through the lenses of security and privacy, however, a package tracking number can possibly reveal certain personal information, leading to security and privacy breaches. In this work, we examine the privacy issues associated with online package tracking systems used in the top three most popular package delivery service providers (FedEx, DHL, and UPS) in the world and found that those websites inadvertently leak users' personal data with a PTN. Moreover, we discovered that PTNs are highly structured and predictable. Therefore, customers' personal data can be massively collected via PTN enumeration attacks. We analyzed more than one million package tracking records obtained from Fedex, DHL, and UPS, and showed that within 5 attempts, an attacker can efficiently guess more than 90% of PTNs for FedEx and DHL, and close to 50% of PTNs for UPS. In addition, we present two practical attack scenarios: 1) to infer business transactions information and 2) to uniquely identify recipients. Also, we found that more than 109 recipients can be uniquely identified with less than 10 comparisons by linking the PTN information with the online people search service, Whitepages.

CCS CONCEPTS

• **Security and privacy → Web application security**; **Privacy protections**; **Data anonymization and sanitization**.

KEYWORDS

Package Tracking, User Privacy, Enumeration Attacks

ACM Reference Format:
Simon S. Woo*, Hanbin Jang*, Woojoong Ji, Hyoungshick Kim†. 2020. I've Got Your Packages: Harvesting Customers' Delivery Order Information using Package Tracking Number Enumeration Attacks. In *Proceedings of The Web Conference 2020 (WWW '20), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3366423.3380062

1 INTRODUCTION

Due to the proliferation of e-commerce businesses and advancements in online shopping, thanks to e-commerce giants, a demand for package shipping services has been skyrocketing. According to a recent report [13], the global parcels market value was almost $380 billion dollars in 2018, and is expected to grow continuously. When shipping packages, typically, all package delivery service providers generate a *package tracking number* (*PTN*) so that the package can be easily tracked and monitored while in-transit, without explicitly logging in to their websites. The FedEx website (https://www.fedex.com/en-gb/tracking.html) explains its own package tracking system as follows:

> *FedEx Tracking, the tool at the top of your page gives you shipment updates in just one click. A tool built for speed, simplicity and convenience. Save your shipments straight to our homepage, so you don't have to log in to find out where your packages are.*

Although the detailed format of tracking numbers vary from one provider to another, a PTN typically consists of 10-18 alphanumeric characters. The tracking number is supposed to be unique per shipment and no two different shipping transactions can have the same tracking number. Also, the PTN should only be known among the sender, the receiver, and the delivery service provider. With a PTN, the package tracking service websites typically display the detailed package status along with some types of personal information, such as the full or partial name of the sender or the receiver, time-stamps, the transit location, the expected delivery time, etc. Indeed, such package tracking systems using PTN seem to be a convenient way to monitor and track the up-to-date package status. Nowadays, most of the package delivery service providers provide a web interface for customers to track their package status on the Internet by simply entering a PTN. However, recent incidents [3, 4] on package delivery services raised security and privacy concerns. Our work is motivated by the interest of analyzing the security and privacy risks of packaging tracking systems. We specifically examine the following research questions: (1) How difficult is it to predict and enumerate PTNs? (2) And what types of personal information can be found in package tracking systems and do they cause exposure to privacy risk?

To address these research questions, we perform a large-scale empirical study to collect and analyze PTNs from the top three package delivery service providers in the world (FedEx [7], DHL [6], and UPS [10]), having market shares of 84% in total [16]. With a few initial samples of PTNs used in FedEx, DHL and UPS, we develop a systematic procedure to predict PTNs from each company.

*First author contributed equally, †Corresponding author.

WWW '20, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7023-3/20/04.
https://doi.org/10.1145/3366423.3380062

Our experimental results demonstrate that real PTNs are not random enough, but highly structured and predictable. In addition, we analyze and compare the extent and the amount of personal information displayed from each package tracking system. Interestingly, we find that FedEx and DHL reveal more personal information than UPS, possibly leading to more personal information leaks. Another serious issue we found is that we can identify 109 individuals with high probability, when this information is combined with other online sources such as Whitepages [11].The contributions of our work are summarized as follows:

- We present a framework to analyze security and privacy issues of package tracking systems using PTN. We specifically develop a kind of dictionary attack called *enumeration attack* to generate customers' PTNs in an automated manner.
- We conduct a large scale study with data from the top three package delivery service providers (FedEx, DHL and UPS) to detect their potential security problems, showing that customers' personal information (e.g., name, home address, etc.) can be harvested by a web crawler.
- Our work is the first to uncover how PTNs can be used to leak users' personal information, and to demonstrate that existing PTN systems can be abused to jeopardize user privacy.

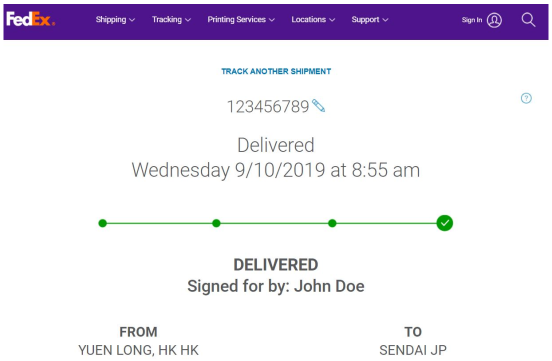
2 OVERVIEW OF PTN SYSTEM

Online PTN System: As shown in Fig. **??**, if a user simply enters a PTN (e.g., '123456789') without logging in, the website displays the package status information for the corresponding tracking number. Interestingly, additional user authentication processes are not required for tracking the user's package delivery status. We surmise that a typical package tracking system is designed to allow even non-members of the system to track the package status. Obviously, this feature is useful for shipping companies, because the sender and/or the recipient can be a non-member of the system who cannot log in to the website. However, this feature may be a double-edged sword, because such online package tracking systems can be abused to harvest users' personal data if we fail to protect PTNs from guessing. In this paper, we analyze how difficult it is to guess other users' PTNs in real-world online package tracking systems.

Structures of PTN: Universal Postal Union (UPU) [9], a specialized agency of the United Nations for regulating the postal service, suggests the UPU standard for postal item identifiers to provide a standard format for PTNs to ensure exchangeability with other shipping companies. However, actual formats of PTNs used in most shipping companies (e.g., FedEx, UPS, and DHL) vary and they are different from the standard format. For example, FedEx uses 12 to 14-digit numbers and UPS uses at most 18 alphanumeric characters, while DHL uses 10-digit numbers. They are summarized in Table 1.

Table 1: PTN formats and maximum spaces for FedEx, UPS, and DHL.

Company	PTN format	Example	Max. space
DHL	10-digits	1332786770	10^{10}
FedEx	12–14 digits	101046952209	10^{14}
UPS	up to 18-alphanum.	1Z83435Y6854645470	36^{18}



TRACK ANOTHER SHIPMENT

123456789

Delivered
Wednesday 9/10/2019 at 8:55 am

DELIVERED
Signed for by: John Doe

FROM
YUEN LONG, HK HK

TO
SENDAI JP

Figure 1: Example of displayed information from FedEx. For illustration, we use "John Doe" as the recipient's name and a fake PTN ('123456789') information.

At first glance, the theoretically available space of 10 or 14-digit numbers, and 18 alphanumeric characters seems sufficiently large to resist against guessing attacks. If an attacker can access a package tracking website (see Fig. **??**), the attacker can try to guess a victim's PTN by enumerating through every possible combination of PTN until a valid one is found. In theory, the attacker needs to attempt a vast amount of guesses (e.g., 10^{10} for DHL), as shown in Table 1. However, in reality, this is not the case, because the actual PTN spaces are much smaller, making guessing attacks feasible.

Displayed personal information with PTN: Next, we analyzed what types of personal information are displayed by PTNs from popular package delivery service providers' (e.g., FedEx, DHL, and UPS) websites. The displayed information can be categorized as follows: 1) the status information, which is the information about whether a package is delivered, in transit, etc., 2) the sender's address, 3) the recipient's name, 4) the recipient's address, and 5) the delivered time of package. FedEx and DHL present information about the sender's city, the recipient's name and city, and the information on time as shown in Fig. 1. On the other hand, UPS shows all of those information except the sender's city. As shown in Fig. 1, however, these first or last names, when combined with a location (city) information, might be sensitive information depending on the context, and may be linkable to other personal information. We found that some types of personal information leaked from PTN systems can be abused for targeted attacks (e.g., [2]). Moreover, due to the rise of the EU General Data Protection Regulation (GDPR) [8] compliance requirements for the protection of users' online privacy, the amount of personal information managed by web services, such as the online PTN system, should be carefully examined and controlled.

3 DATA COLLECTION VIA PTN ENUMERATION ATTACKS

In this section, we analyze how secure current PTN-based package tracking systems are against enumeration attacks. We propose an enumeration attack [12, 14, 15, 17] for PTNs, PTN guessing attack, which is a type of dictionary attack in which an attacker tries each PTN from a list of possible candidate PTN values. The correctness of

Table 2: Information displayed from online package tracking number status services.

Company	Status Info.	Sender Addr.	Recipient Info.	Recipient Addr.	Delivered Date	Additional Info.
FedEx	Yes	City, Country	Last name, Initial of First name	City, Country	YY-DD-MM-HH-MM	–
DHL	Yes	City, Country	Last or First name	City, Country	YY-DD-MM	–
UPS	Yes	–	Last or First name	City, Country	YY-DD-MM-HH	Type, Weight, Left at

Table 3: Collected Dataset Description (Note: UPS does not provide sender information).

Company	Total num. of live PTNs	Num. of unique sender city/country	Num. of unique recipient city	Num. of unique (recipient name,recipient city)	Num. of unique (sender city, recipient name, recipient city)
FedEx	101,493	6,496	17,885	38,367	44,164
DHL	451,457	19,530	50,586	278,192	337,160
UPS	379,627	N/A	39,281	240,604	N/A

the tried PTN value is verified through the received online feedback. We note that PTN enumeration attacks, guessing attempts, can be repeated, until the correct *secret* (e.g., PTN) value is found. The main objective of PTN enumeration attacks is to collect additional PTNs, which can help discover the underlying patterns and predict future PTNs. Based on a statistical analysis with the collected PTNs, we develop novel enumeration attacks that are highly effective at harvesting users' personal information (e.g., recipient's name and address). The details of our enumeration attacks are presented in the following section.

3.1 Data Collection Methodology

We first manually collected a small number of PTNs used in FedEx, UPS and DHL, and analyzed them to identify the structure and format of PTNs for each system. Next, we developed an enumeration attack, exploiting the detailed structure of PTNs to harvest PTNs in an automated manner. In all the package tracking services that we examined, we found that there is no security policy to limit the number of consecutive attempts.

For FedEx and DHL, we performed enumeration attacks to directly harvest PTNs. For UPS, on the other hand, we used a different approach to harvest PTNs, because UPS has a large theoretical PTN space of 36^{18} (see Table 1). We discovered that UPS has used another package tracking option by assigning a customized *reference number* to each package, which makes it easier to correlate with the customers' information. Hence, for UPS, reference numbers can be alternatively used, instead of PTNs, to track the package delivery status [1]. Interestingly, most reference numbers that we observed were relatively short (around 18 characters), even though a reference number can have up to 35 characters at maximum. Thus, for UPS, it is even easier to enumerate reference numbers than PTNs. Also, the use of reference numbers is helpful to speed up the process of enumeration attacks, because a reference number is associated with multiple PTNs (about 45 PTNs on average in our experiments).

Initial seed PTNs for FedEx and DHL: Initial seed PTNs can be easily obtained both online and offline to start an enumeration process. We used PTNs to order our own packages or to send them. Also, we used a search engine to find some initial PTNs. For FedEx, we obtained 18 PTNs from various geographical locations and used them as seed PTNs. They include packages sent and received from China, USA, Greece, and South Korea. Therefore, we have data corresponding to geographically diverse senders across the globe. For DHL, we obtained 51 PTNs from various Asian and European countries through Internet search. With the initial seed PTNs, we analyzed the valid format of PTNs used for each tracking system. Next, we generated candidate PTNs within a range of numbers (from an initial seed PTN) via PTN enumeration attack by adding certain numbers, and tried to search for tracking information with the candidate PTN on the tracking service website. If the package tracking information was successfully returned from the website, the information was crawled and stored in a database; otherwise, we sequentially repeated the searching and crawling steps using the next candidate PTN, where these steps can be automatically processed. In order not to interfere with the normal operation in providers' websites, we queried the websites' tracking service at a very slow rate.

Collecting initial seed PTNs from UPS: For UPS, we use reference numbers instead of PTNs, which are highly similar to PTNs but more efficient. We first collected around 10 PTNs from public websites and obtained a few reference numbers associated with those PTNs. We used those reference numbers as the initial seeds for enumeration attacks to collect reference numbers from the tracking service website for UPS. The process of enumeration attacks on UPS was identical to those on FedEx or DHL, except for the fact that reference numbers were used instead of PTNs. With the generated reference number candidates, we crawled the package tracking information as well as the corresponding PTNs associated with valid reference numbers obtained through our PTN enumeration attacks.

3.2 Statistics of Collected Data

We collected 101,493 live PTNs for FedEx from Oct. 2017 to May 2019, 451,457 live PTNs for DHL from Feb. 2019 to June 2019, and 379,627 live PTNs for UPS from Feb. 2019 to July 2019. At the time of data collection, we were able to collect more PTNs for DHL and UPS than for FedEx, because of the space and time constraints during our experiment. However, we collected datasets of at least 100,000 live PTNs for each provider to demonstrate the issues regarding PTNs. All the high level statistics are summarized in Table 3. In total, we collected data corresponding to 19,530 sender cities for DHL, and 6,496 cities for FedEx over 191 countries. Since UPS does not provide the sender's city information, we do not present it. In addition, we collected data corresponding to 50,586, 17,885, and 39,281 unique recipient cities for DHL, FedEx, and UPS, respectively.

Taking into account unique recipient names and city information together (e.g. John Smith in NYC, NY vs. John Smith in Eureka, CA), we found that there are 38,367, 278,192, and 240,604 unique pair information for FedEx, DHL, and UPS, respectively. Further, we consider the sender's city, the recipient's city and the recipient's name together (e.g., a package sent from "Berlin" to "John Smith" in "Eureka, CA" vs. a package sent from "Shanghai" to "John Smith" in "Eureka, CA").

4 ANALYSIS OF PTN DATA

We aim to predict the next PTN from the known PTNs, using the underlying patterns between PTNs. Therefore, to examine patterns in a sequence of PTNs, we first measure the difference between two consecutive PTNs, $\Delta PTN(i)$=$PTN(i+1)$-$PTN(i)$, where $PTN(i)$ is the ith PTN in the sequence of PTNs. We calculate the cumulative distribution functions (CDFs) of $\Delta PTN(i)$ (from 1 to 10^4) for all the collected PTNs from FedEx, DHL, and UPS tracking systems, respectively, to obtain the differences between two consecutive PTNs. Hereafter, we denote $\Delta PTN(i)$ for all i in the collected PTNs as ΔPTN. Figure 2 shows the calculated CDFs for FedEx, DHL, and UPS, where the X-axis represents the difference between two consecutive PTNs in $\log_{10}$, and the Y-axis represents the cumulative percentage of the number of PTNs. Similarly, for UPS, PTNs consist of alphanumeric charterers. Hence, we calculate ΔPTN as the difference between the corresponding base-36 numbers.

Figure 2 shows an upward trend of the cumulative percentage of the number of PTNs. For certain points in ΔPTN, we observe a high spike with a stepwise pattern, indicating that the differences between two consecutive PTNs are highly concentrated at those points. For example, when ΔPTN=1 ($\log_{10}(1)$ = 0), we observe an increase in the Y-axis of around 7% for FedEx (shown as a blue line in Figure 2). When ΔPTN=10 ($\log_{10}(10)$ = 1), we observe an even larger increase of around 60%. This indicates that if we add 1 or 10 to PTNs, we can obtain 7% or 60% chance of getting the PTNs. These stepwise increase patterns are also present for DHL and UPS (shown as orange and yellow lines, respectively, in Fig. 2). For example, we observe an increase in the Y-axis of around 70% for DHL, when ΔPTN=11 ($\log_{10}(11) \approx 1$). This means that if we add 11, we can obtain another PTN with a 70% chance. For UPS, we first observe a large increase (33%) in the Y-axis for $\log_{10}(35) \approx 1.54$, when ΔPTN=35. Similary, this means that if we add 35 to the valid DHL PTN, we can get the next PTNs with a 33% chance. This shows that there exist underlying patterns, and adding specific ΔPTNs can be more effective in predicting and obtaining PTNs.

In Table 4, we tabulate the top 5 ΔPTNs yielding the largest increase in CDF for each service. For FedEx, $\Delta PTNs$ of 11, 10, 1, 15, and 22 shows the largest increase, constituting 90.9% of the total PTNs, as shown in Table 4. This means that if we add each of those five numbers sequentially to a FedEx PTN, we can obtain the next PTN with about 90.9% chance. For DHL, a ΔPTN of 11 yields the highest increase with 75.2%, as shown in Table 4. The second and third highest ΔPTNs are 4 and 48 with a 9.9% and a 7.8% increase, respectively. Therefore, if we add each of those five numbers sequentially to a DHL PTN, we can obtain the next PTN with about 98.3% chance. For UPS, we can also find a similar trend, even though the probability of finding the next PTN becomes relatively smaller — the top 5 ΔPTNs take about 48.9% as shown in Table 4. We surmise that this is because UPS uses a longer and more complex PTN format (see Table 1), compared to FedEx and DHL.

Table 4: Top 5 ΔPTN values with the highest proportion of predicting PTNs and 'Total' is the sum of top 5 ΔPTN values (%).

Provider	Top 5 ΔPTN values (%)					Total
	Top 1	Top 2	Top 3	Top 4	Top 5	
FedEx	11 (68.5%)	10 (7.5%)	1 (7.5%)	15 (4.2%)	22 (3.2%)	90.9%
DHL	11 (75.2%)	4 (9.9%)	48 (7.8%)	70 (4.5%)	22 (0.9%)	98.3%
UPS	35 (33.0%)	70 (7.0%)	45 (3.7%)	105 (2.6%)	969 (2.6%)	48.9%

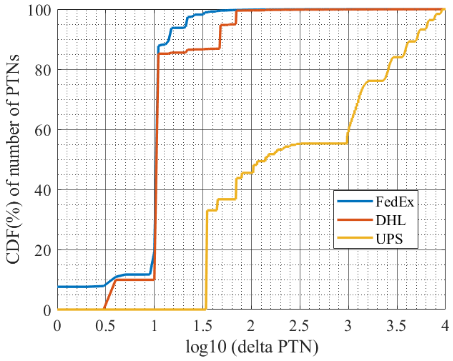


Figure 2: Cumulative distribution functions (CDFs) of ΔPTN for FedEx, DHL, and UPS, where the X-axis is in $\log_{10}$ and the Y-axis is percentage (%).

Our analysis indicates that within 5 attempts, an attacker can guess more than 90% of PTNs for FedEx and DHL, and close to 50% of PTNs for UPS, using the patterns of consecutive PTNs as shown in Table 4. Moreover, as shown in Fig. 2, we can obtain the next PTN from known PTNs within 10,000 guesses for all services.

5 ATTACK SCENARIOS

In this section, we present two possible attack scenarios, where the collected delivery order information can be abused to infer sensitive and private user or company proprietary information.

5.1 Attack 1. Inferring business and customer information

From the collected package delivery data as shown in Section 3.2, we found that a few senders have dominantly placed a large volume of orders. Perhaps, they are not normal customers, but could be suppliers. Attackers may be able to track such suppliers' delivery orders to infer their business transactions and (regular) customers. Surely, such business transaction information (e.g., who bought and where it shipped) has to be securely protected, because they are often treated and can be regarded as company proprietary information.

Experiment: As a case study, we ordered a drone from DJI, which is one of the largest drone manufacturers in the world, and they used FedEx as the shipping company. As a result, we were able to obtain a PTN from our shipping order and used it as a seed PTN to collect other PTNs, using the enumeration attacks described in Section 3. From the collected PTNs, we also collected the delivery order data associated with those PTNs and filtered them by their sender address, which is the same as that of the seed PTN to selectively collect the delivery order data related to DJI. Finally, we collected 1,412 delivery orders, which were used to deliver their packages to customers from a seed PTN.

Result: From the obtained information, we are able to specifically analyze and identify where, when, and to whom DJI is sending its products using FedEx. We visualize the delivery order relationships between cities in the world map (see Fig. 3). The shipping location is Hangzhou, and their products are delivered to customers all over the world as shown in Fig. 3 and we were able to track every business and customer transactions.

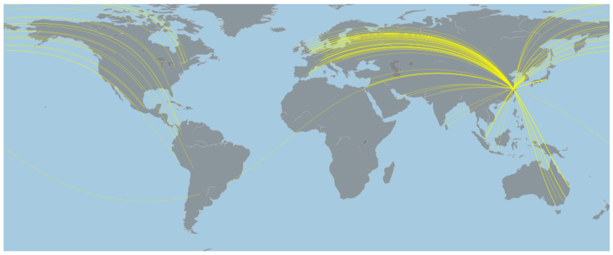


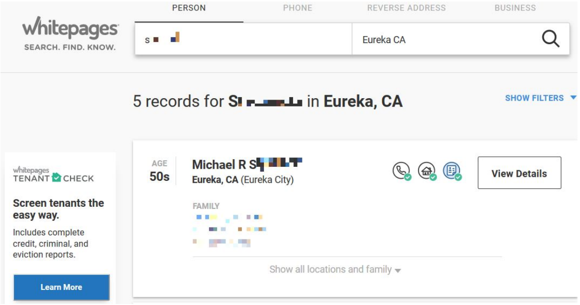
Figure 3: Delivery order relationships using FedEx between DJI and its customers.

Summary: This case study shows that the collected PTNs can be used by attackers to extract the business information (customer info, destination, etc.) of a target company. For example, competitors can easily obtain information about a target company in the same way we proposed. Moreover, we believe that companies have a responsibility to protect their customers' personal information, including their location and name. Hence, a data leakage through flaws in online PTN systems can damage a company's reputation and brand name, and compromise the privacy of their customers.

5.2 Attack 2. Identifying individuals using Whitepages

As a second case study, we show that the collected delivery order data can be used by attackers to identify individuals, using online people search services such as Whitepages [11], which is one of the largest online and offline directory service provider in the USA.

Whitepages provides several options to search people on the database (e.g., by name, city, state, ZIP or phone number). For example, when a target person's name and city are typed into Whitepages search bar, the results from people with the same name and residing city will be displayed. Also, it displays people's age range information, as well as their current and past address information in detail. Fig. 4 shows an example of a search result returned (5 users) from Whitepages with a query name = "S****" and location = "Eureka, CA."



s [redacted]

Eureka CA

5 records for S[redacted] in Eureka, CA

Screen tenants the easy way.

Includes complete credit, criminal, and eviction reports.

AGE 50s

Michael R S[redacted]

Eureka, CA (Eureka City)

FAMILY

[redacted]

Figure 4: Example from Whitepages — for privacy reasons, the details of the names are blurred.

Experiment: In this experiment, we only considered recipients who live in the USA, because Whitepages only works for USA residents. To show the feasibility of this attack scenario, we randomly selected 317 unique recipients who have completed delivery transactions and have a clear name from the UPS delivery order data. We searched for people with those names and cities. For each search query, we counted the number of people displayed by Whitepages (i.e., the number of people who have the same name and reside in the same city).

Result: Through 317 queries on Whitepages, we collected 1,618 people records. From these obtained results, we found that the identities of about 34.4% of recipients can be narrowed down to fewer than 10 individuals. For example, only 5 inspections are needed to identify the queried individual in Eureka. Therefore, 1,618 people are identifiable with less than 600 comparisons and 317 recipients returned less than 100 records from Whitepages. Surprisingly, 15 people (4.7%) are uniquely identifiable on Whitepages with only one returned result. 18 query results return 2 people records, which means we only need to inspect 2 people to identify the recipient, assuming the age ranges and full records of their current and past addresses provided by Whitepages. Therefore, if we use PTNs and Whitepages together, we can fully identify a significant number of individuals with high probability.

Summary: Our result shows that recipients are not only linkable but also identifiable with high probability (100% for 15 users), when combined with Whitepages. This clearly jeopardizes the privacy of an individual. Furthermore, our result shows how easily additional personal information (full and past addresses, age range, etc.) can be found, when combined with other online information such as Whitepages. Although we manually implemented this attack to comply with Whitepages Terms of Service (https://www.whitepages.com/terms-of-service), we expect that the entire process of the attack can be fully automated. Therefore, it would be urgent to fix the flaws in package tracking systems that allow attackers to collect PTNs and use them to harvest people's personal information from various sources such as Whitepages.

6 DEFENSE MECHANISMS

To address the problem of PTN enumeration attacks, one of the simplest, yet effective ways is to enhance the randomness of PTNs, making PTNs harder to enumerate within a fairy short amount of time.

However, current PTNs cannot be fully randomized, because some digits in a PTN represent specific semantic information (e.g., location and time) about package delivery orders, which are often hardcoded. Another promising approach is to employ CAPTCHAs [19], which can block automated attempts. Surprisingly, we found that none of UPS, DHL, or FedEx employs CAPTCHAs. Therefore, massive information exposure and leaks are possible. However, it is not easy to deploy both secure and usable CAPTCHAs, because the computational gap between a human and AI may not be significantly large and computer vision and machine learning algorithms have shown to bypass some of the CAPTCHAs easily [21]. Another possible defense approach is to employ a stronger authentication process, such as the multi-factor authentication, so that authorized parties (e.g., sender and recipient) can access their package delivery order data only after verifying themselves with additional secrets. Although this may cause additional inconvenience and changes in the PTN tracking system, it can prevent strangers on the Internet from easily accessing other people's package status information. For usability, we can simply check whether a verifier knows the recipient's phone number or email address instead of using a complex user authentication scheme. Also, to mitigate such automated enumeration attacks, we also recommend using a security policy to limit the number of PTN query attempts from a particular user such as IP addressess. When personal information needs to be displayed publicly online, it has to be properly anonymized to protect user privacy. For example, masking some parts of the customer name in a privacy-preserving manner may hinder the process of identifying specific individuals.

7 ETHICAL ISSUES

We addressed the ethical concerns of our study as follows: First, we gained an IRB-approval from our institution to conduct this research. Second, we designed experiments in such a way to avoid interfering with the normal operations of online PTN-checking service by querying at a slow rate (around 1 sec. interval). We performed the t-test for the website response time with and without our query to show there is no significant statistical difference (p-values for FedEx, UPS, and DHL are 0.745, 0.907, and 0.447, respectively). Lastly, all collected data are publicly available on the Internet and are openly accessible using PTNs.

8 RELATED WORK

As the package delivery industry becomes more popular, the customer data on online package delivery systems would become an increasingly attractive target for attackers. According to the recent report by Shoorbajee [18], TNT Express, FedEx's Dutch subsidiary, attributed an estimated $300 million loss to the NotPetya ransomware that impacted the company's operations. In 2018, Kromtech, a security research company, reported that FedEx was hacked, indicating that more than 119,000 FedEx customers' scanned documents including passports, driver's licenses and security identification were leaked [3]. Similarly, USPS experienced a data breach, which may have resulted in attackers compromising more than 60 million customers' personal information and online account details including email addresses, user names, phone numbers and home address [4]. Canada Post also experienced a data breach in 2018 [20] — an outsider gained accessed to data of about 4,500 customers through Canada Post's delivery tracking tool, exposing sensitive delivery information such as postal codes, names or initials of the people who signed for the order, corporate names, business addresses, PTNs, etc. Although all shipping companies claimed to fix their data breach problem after they have experienced it, our experimental results demonstrate that their package tracking systems still pose a serious problem, which could expose customers to several security and privacy threats. To show the potential security and privacy risk in online package tracking services, we develop an enumeration attack. Potential security risks of enumeration attacks have been recently studied in instant messaging services and social network services (SNS). Schrittwieser et al. [17] presented an enumeration attack to collect 21,095 live phone numbers from WhatsApp [5] within less than 2.5 hours. Kim et al. [14] also collected 50,567 users' personal data such as phone number, displayed name, and profile picture from KakaoTalk, which is the most widely used instant messaging service in South Korea, through enumeration attacks. Balduzzi et al. [12] collected 10.4 million e-mail addresses by using Facebook's 'friend finder'. Kim et al. [15] also presented a new enumeration attack by mimicking multiple users' search activities with a few *Sybil* accounts and collected 82,082 Facebook users' personal data in 15 days. Unlike SNS data, however, our work is the first to systemically analyze the possibility of enumeration attacks and the privacy leakage problem tailed to the package tracking services.

9 CONCLUSION

In this work, we shed light on how PTNs that are used on a daily basis can be exploited personal information leakage. We found that PTNs can easily be predicted and further used to mine customers' personal information. Our experimental results show that top shipping companies did not fully consider a reasonable level of security practices to protect the customers' personal information stored in their servers. We also introduce two realistic attack scenarios demonstrating how collected user data can be abused. Most notably, we were able to fully identify 15 individuals living in the USA, using Whitepages. This means that these people's current and past addresses, as well as age ranges, are fully disclosed. Since this issue requires an immediate attention to minimize personal information leakage, we strongly hope that companies would take immediate actions. For future work, we plan to collect seed PTNs from different geographical regions and conduct further analysis with various service providers.

ACKNOWLEDGEMENTS

This research was supported by Energy Cloud R&D Program through the National Research Foundation (NRF) of Korea funded by the Ministry of Science, ICT (No. 2019M3F2A1072217), the Basic Science Research Program through the NRF of Korea funded by the Ministry of Science, ICT (No. M2017R1C1B5076474, NRF-2019R1C1C1007118), the ICT Consilience Creative program (IITP-2019-2015-0-00742), and by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-00421, AI Graduate School Support Program).

REFERENCES

[1] 2011. Reference Numbers. https://www.ups.com/worldshiphelp/WS14/ENU/AppHelp/GlossPopup/Reference_Numbers.htm Accessed: 2019-10-10.

[2] 2014. Cybercriminals Use What Works (Targeted Attack Methodologies for Cybercrime). https://www.trendmicro.co.uk/media/wp/cybercriminals-use-what-works-whitepaper-uk.pdf Accessed: 2019-10-08.

[3] 2018. FedEx Data Breach. https://www.informationsecuritybuzz.com/expert-comments/fedex-data-breach/ Accessed: 2019-10-14.

[4] 2018. USPS Site Exposed Data on 60 Million Users. https://krebsonsecurity.com/2018/11/usps-site-exposed-data-on-60-million-users/ Accessed: 2019-10-14.

[5] 2019. About WhatsApp. https://www.whatsapp.com/about/ Accessed: 2019-10-10.

[6] 2019. DHL Global. http://www.dhl.com/en.html Accessed: 2019-10-14.

[7] 2019. FedEx. https://www.fedex.com Accessed: 2019-10-14.

[8] 2019. General Data Protection Regulation (GDPR). https://eugdpr.org/ Accessed: 2019-10-14.

[9] 2019. Universal Postal Union. http://www.upu.int/en.html Accessed: 2019-10-14.

[10] 2019. UPS. https://www.ups.com Accessed: 2019-10-14.

[11] 2019. Whitepages – Find people, contact info & background checks. https://www.whitepages.com/ Accessed: 2019-09-10.

[12] Marco Balduzzi, Christian Platzer, Thorsten Holz, Engin Kirda, Davide Balzarotti, and Christopher Kruegel. 2010. Abusing Social Networks for Automated User Profiling. In *Proceedings of the 13th International Conference on Recent Advances in Intrusion Detection*. Springer-Verlag, 422–441.

[13] Apex Insight. 2019. Global Parcel Delivery Market Insight Report. https://apex-insight.com/product/global-parcel-delivery-market/ Accessed: 2019-10-14.

[14] Eunhyun Kim, Kyungwon Park, Hyoungshick Kim, and Jaeseung Song. 2015. Design and analysis of enumeration attacks on finding friends with phone numbers: A case study with KakaoTalk. *Computers & Security* 52 (2015), 267–275.

[15] Jinwoo Kim, Kuyju Kim, Junsung Cho, Hyoungshick Kim, and Sebastian Schrittwieser. 2017. Hello, Facebook! Here is the stalkers’ paradise!: Design and analysis of enumeration attack using phone numbers on Facebook. In *Proceedings of the 13th International Conference on Information Security Practice and Experience*.

[16] E. Mazareanu. 2019. Couriers and local delivery service providers’ global market share in 2017. https://www.statista.com/statistics/236309/market-share-of-global-express-industry/ Accessed: 2019-10-14.

[17] Sebastian Schrittwieser, Peter Frühwirt, Peter Kieseberg, Manuel Leithner, Martin Mulazzani, Markus Huber, and Edgar R. Weippl. 2012. Guess Who’s Texting You? Evaluating the Security of Smartphone Messaging Applications. In *NDSS*. The Internet Society.

[18] Zaid Shoorbajee. 2017. FedEx attributes $300 million loss to NotPetya ransomware attack. https://www.cyberscoop.com/fedex-attributes-300-million-loss-notpetya-attack/ Accessed: 2019-10-14.

[19] Luis Von Ahn, Manuel Blum, Nicholas J Hopper, and John Langford. 2003. CAPTCHA: Using hard AI problems for security. In *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 294–311.

[20] Oriena Vuong. 2018. Canada Post data breach affected 4,500 customers, OCS says. https://energy953radio.ca/news/4639742/ocs-canada-post-hacked/ Accessed: 2019-10-14.

[21] Guixin Ye, Zhanyong Tang, Dingyi Fang, Zhanxing Zhu, Yansong Feng, Pengfei Xu, Xiaojiang Chen, and Zheng Wang. 2018. Yet another text captcha solver: A generative adversarial network based approach. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. 332–348.