Your Purchase History Has Been Exposed: Privacy Threats on Auction Sites

Ayako Hasegawa
NTT Secure Platform Laboratories
3-9-11 Midori-Cho,
Musashino-Shi, Tokyo, Japan
hasegawa.ayako@lab.ntt.co.jp

Tatsuya Mori
Waseda University
3-4-1 Okubo, Shinjuku, Tokyo, Japan
mori@nsl.cs.waseda.ac.jp

Mitsuaki Akiyama
NTT Secure Platform Laboratories
3-9-11 Midori-Cho,
Musashino-Shi, Tokyo, Japan
akiyama.mitsuaki@lab.ntt.co.jp
akiyama@ieee.org

Takeshi Yagi
NTT Secure Platform Laboratories
3-9-11 Midori-Cho,
Musashino-Shi, Tokyo, Japan
yagi.takeshi@lab.ntt.co.jp

Paste the appropriate copyright statement here. ACM now supports three different copyright statements:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single spaced in a sans-serif 7 point font.
Every submission will be assigned their own unique DOI string to be included here.

Abstract

An online purchase history contains a wealth of privacy-sensitive information. To protect buyers from the leakage of their purchase histories, online auction sites such as eBay have adopted some form of privacy protection mechanisms to thwart the threats of unexpected data collation. However, as Minkus et al. [2] have demonstrated, it is possible on eBay to reconstruct an online purchase history by carefully collating the partially masked user IDs and the timing of bidding, which are both available on its rating system. In this study, we extend their work and demonstrate that a purchase history reconstruction attack can also work for other auction sites with more powerful privacy-protection mechanisms, such as, incomplete links between buyer and seller, full randomization of user IDs, and coarse-grained time information. This fact implies that online auction sites need a rating system that can completely protect users' privacy without sacrificing core functionality. Additionally, we study users' expectations regarding their privacy on online auction sites.

Author Keywords

information leakage; purchase history; auction sites;

ACM Classification Keywords

K.4.4 [COMPUTERS AND SOCIETY]: Electronic Commerce

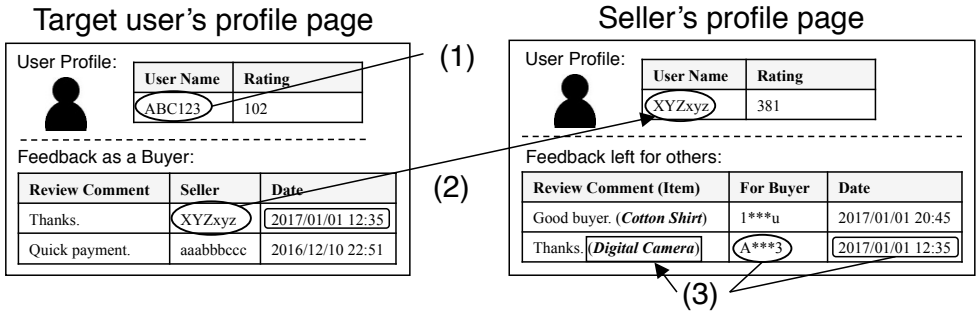

Figure 1: Diagram of the purchase history reconstruction attack.

Table 1: Specifications of eBay and Yahoo! Auction.

Publicly available feedback published on a buyer/seller's profile pages

	eBay	Yahoo! Auction
As a buyer	√	√
As a seller	√	√
Left for others	√	n/a

Publicly available information published on a buyer's profile page

	eBay	Yahoo! Auction
Bidding time	√	n/a
Purchased items	n/a	n/a
Seller's IDs	√	√

Publicly available information published on a seller's profile page

	eBay	Yahoo! Auction
Bidding time	√	n/a
Sold items	√	√
Masked buyers' IDs	Predic-table	Unpred-ictable

Introduction

The use of online auction services has increased in popularity. eBay, one of the most popular online auction websites in the world, reached 169 million active users in 2017 [3]. Online auction sites usually provide a rating system that enables a buyer/seller to determine whether the counterparty is trustworthy by assessing ratings that are built based on past transactions. The rating system typically involves gathering records, each with the following information: a purchased item, an anonymized buyer, a seller, and the ratings given to the seller and/or buyer.

Since collating these data could lead to the exposure of an individual's purchase history, the online auction sites have adopted various forms of privacy-protection mechanisms (e.g., user ID obfuscation) without sacrificing the usefulness of the rating systems. However, in 2014, Minkus et al. [2] clarified potential privacy defects in the eBay's rating system. They exploited the publicly available data obtained from the feedback system and succeeded in identifying some target users' purchases. They also studied the privacy awareness of eBay users and discovered that only 9% of users correctly recognized that their purchases are available for anyone, even if they are not signed in.

In the present study, we extend the work of Minkus et al. [2]. Our research question is as follows:
Can we reconstruct a purchase history on an online auction site that uses powerful privacy protecting mechanisms?
As the first step toward answering the research question, we analyzed data collected from the largest online auction site in Japan, Yahoo! Auction. This auction site adopts a powerful privacy-protection mechanism that includes incomplete links between a buyer and seller, full randomization of user identities, and coarse-grained time information. As a complement to the above study, we also make a repetition of the user study performed by Minkus et al.[2] to understand the privacy awareness of online auction users. We report similarities and differences between their findings and ours.

Purchase History Reconstruction Attack

In this section, we show the key elements of a purchase history reconstruction attack through the example proposed by Minkus et al. [2]. Figure 1 presents a diagram of a purchase history reconstruction attack on eBay. In general, items purchased by a buyer are not directly associated with his or her profile page. However, an attacker can identify the buyer's purchase through the following simple process. (1) Given the profile page of a buyer (`ABC123`), (2) an attacker can track the profile page of a seller (`XYZxyz`) who rated the buyer. (3) On the seller's profile page, the attacker can find the purchased item was likely a digital camera from the corresponding bidding time (Jan 1, 2017 12:35). The attacker can also ensure that the estimate is correct because the masked ID, "`A***3`", has two letters that also reside in the original ID "`ABC123`" as the convention includes two common letters in the strings. The key success factors

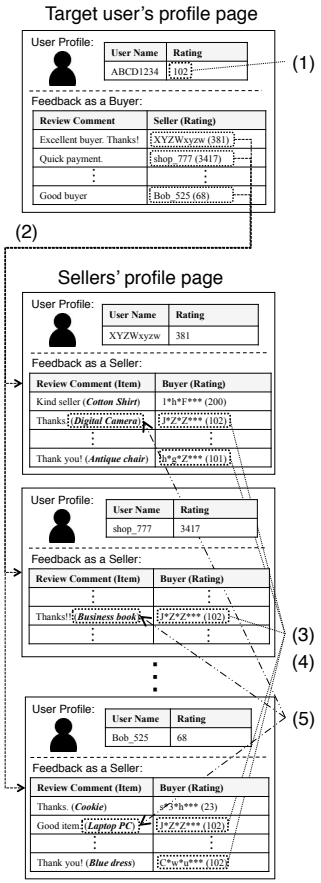

Figure 2: Diagram of the extended attack.

of this attack are 1) the existence of a clear link between the seller and buyer and 2) explicit matching of the bidding times. [1]

Extended Attack

We develop an extended attack that can identify a user's purchases for auction sites that have a more powerful privacy protection scheme than eBay. As an example of such auction sites, we targeted Yahoo! Auction [1]. Table 1 lists the differences between the specifications of Yahoo! Auction and eBay. The chief differences are as follows:

An incomplete link between sellers and buyers: While a link from buyer to seller is automatically generated when every user rates the counter-party in eBay, that link is generated when a buyer rates a seller in Yahoo! Auction. Therefore, tracking the purchases from a given buyer's page is impossible if the buyer does not rate sellers.

No bidding time: While eBay has adopted the time stamp for presenting the timing of bidding, Yahoo! Auction adopts the coarse-grained time stamp for the time the rating is received. Therefore, the information about transaction time recorded on a buyer's and a seller's profile pages cannot be collated.

Random pseudonyms: While eBay allows an attacker to check whether the pseudonyms of buyers match the original ID by simply checking the inclusion of unmasked letters, Yahoo! Auction adopts fully random pseudonyms.

To overcome the restrictions shown above, we develop an extended attack that leverages the rating score.

Figure 2 shows the diagram of the extended attack.
(1) Check the rating score of a target buyer (102).
(2) Collect the profile pages of the sellers from whom the target buyer purchased items.
(3) Extract the pseudonyms that have rating scores close to the target buyer's rating score ((`J*Z*Z***`, `h*g*Z***`), (`J*Z*Z***`), $\cdots$, (`J*Z*Z***`,`C*w*u***`)).
(4) Compute the frequencies of the extracted pseudonyms. Extract the pseudonym that had the highest frequency (`J*Z*Z***`).
(5) The extracted pseudonym of the target buyer can be associated with the purchased items (digital camera, business book, and laptop PC).

We assessed how our attack can identify the user purchases through real world data. To this end, we introduced a metric that represents the distinguishability of the extracted candidates; i.e., if many candidate accounts had similar frequencies at step (4), the attack might fail. The distinguishability score is computed as $s = f_1/f_2$, where f_1 and f_2 were the highest frequency and second to the highest frequency, respectievly. If $s \geq 2$, we concluded that the pseudonym with highest frequency was enough to be identified as the target buyer; i.e., we adopted the maximum likelihood estimation approach. We collected 144 Yahoo! Auction accounts that have at least two ratings as a buyer. For each account, we applied the attack and compute the distinguishability score defined above. Figure 3 shows the result. Of the 144 accounts, 137 accounts had the score of $s \geq 2$. That is, 95.1 % of the accounts were vulnerable to the purchase history reconstruction attack. Moreover, we also noticed that as the number of ratings as a buyer increased, the distinguishability also increased.

Finally, to verify the success of attack, we applied the reconstruction attack to the two Yahoo! Auction accounts, which we know the ground truth. For these two accounts, the attack perfectly reconstructed the full purchase history.

[1] As of May 2017, eBay has changed its specifications; the exact matching of bidding time is not applicable for the current system.

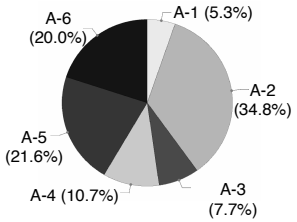

Figure 4: Who can see your purchases?

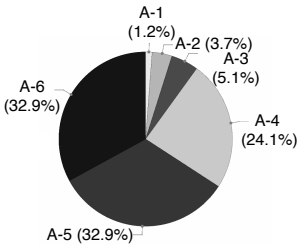

Figure 5: Who can see the feedback left for you?

Answers:
[A-1]: no one
[A-2]: just me
[A-3]: all sellers
[A-4]: the sellers left feedback for me
[A-5]: anyone sighed into auction service
[A-6]: anyone, even not signed into auction service

Both questions, the correct answer is [A-6].

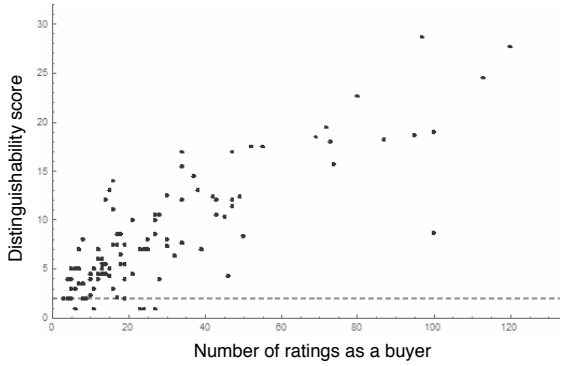

Figure 3: Number of ratings vs. Distinguishability score

User Expectations

Minkus et al. [2] conducted a survey of eBay users to investigate their expectations and behaviors. We conducted the same survey of Yahoo! Auction users using a Japanese crowd sourcing service. We asked 9 muiltiple-choice questions, paying about 0.25 USD for answering each, resulting in 431 valid responses. They found an interesting result that most eBay users are not aware of who can see their purchases. Figure 4 shows the answers to the same question asked to Yahoo! Auction users, "Do you know who can see your purchases?". We found that 20% users answered correctly, compared to only 9% eBay users in the previous study [2]. Additionally, Figure 5 shows the answers to the question, "Do you know who can see the feedback sellers left for you?". Figure 4 and Figure 5 indicate users tended to believe that their purchases were not public even though they recognized that feedback is public.

Discussion

A key finding we derived is that if users often rates each other, it makes easier for an attacker to identify the purchase history of that user. In other words, high activity on an auction site could turn into the footprint of a user's purchase history for an attacker. This finding poses a challenge to the rating system of online auction sites. The following suggestions are possible solutions to this concern in the rating systems. For eBay, its "Private Listing" option[2], which removes any visible links between the buyers and sellers, could be a possible solution. Another is to adopt *non-persistent* pseudonyms for buyers because this practice makes it more difficult for an attacker to search for a link between the rating information and a user. While these solutions are effective in protecting users from the purchase history reconstruction attack, they could sacrifice the *transparency*, which is a crucial feature to make the entire rating system trustworthy. We note that there is no direct solution for the buyers other than using multiple services to distribute the footprints. As our user expectation study indicates, to make sure buyers understand the threats of a purchase history reconstruction attack is a good starting point so that they can evade unnecessary privacy breaches.

Summary and future Work

We demonstrate that a purchase history reconstruction attack can work even for auction sites with highly powerful privacy-protection mechanisms, including incomplete rating links, user ID randomization, and obscuring time information. Understanding the generic trade-offs between the effectiveness of privacy protection and the usefulness of rating systems remains a goal for our future work.

[2] http://pages.ebay.com/help/sell/private.html

REFERENCES

1. Yahoo! Auction. 2017. (2017). https://auctions.yahoo.co.jp/
2. Tehila Minkus and Keith W. Ross. 2014. I Know What You're Buying: Privacy Breaches on eBay. *Privacy Enhancing Technologies, pp 164-183* 18, 4 (2014), 765–766. http://dx.doi.org/10.1007/s00779-014-0773-4
3. Statista. 2017. Number of eBay's active users from 1st quarter 2010 to 1st quarter 2017. (2017).