

I Know What You're Buying: Privacy Breaches on eBay

Tehila Minkus¹ and Keith W. Ross^{1,2}

¹ Dept. of Computer Science and Engineering, NYU

² NYU Shanghai

tehila@nyu.edu, keithwross@nyu.edu

Abstract. eBay is an online marketplace which allows people to easily engage in commerce with one another. Since the market's online nature precludes many physical cues of trust, eBay has instituted a reputation system through which users accumulate ratings based on their transactions. However, the eBay Feedback System as currently implemented has serious privacy flaws. When sellers leave feedback, buyers' purchase histories are exposed through no action of their own. In this paper, we describe and execute a series of attacks, leveraging the feedback system to reveal users' potentially sensitive purchases. As a demonstration, we collect and identify users who have bought gun-related items and sensitive medical tests. We contrast this information leakage with eBay users' privacy expectations as measured by an online survey. Finally, we make recommendations towards better privacy in the eBay feedback system.

1 Introduction

Online commerce has introduced new risks and rewards for consumers. It offers ease and convenience, allowing for in-depth comparison shopping from the comfort of one's home computer or mobile device. However, the impersonal and intangible nature of online transactions gives rise to trust-based issues as well: how can users know that they will actually receive the goods they bought? Will the goods arrive intact and in a timely fashion? In response to these issues, online marketplaces have instituted reputation systems, where parties to the market are rated based on their behavior in transactions.

eBay is somewhat unique among online marketplaces in that its reputation system is symmetric: not only can buyers rate sellers, sellers can also provide feedback on the users who have bought their wares. At first, this seems like a helpful mechanism; users receive recognition for prompt payment, and a sense of reciprocity may motivate them to contribute feedback to their seller in return. This makes the reputation system robust and popular. However, as we will show in this paper, the current implementation has some serious privacy implications.

In this research, we explore the privacy issues that are byproducts of the symmetric and public nature of the eBay feedback system. We first describe the *purchase history attack*: given a user's eBay username, we show how to discover his purchases by correlating his feedback page with the feedback pages of the

sellers with whom he has interacted. If the attacker knows the real identity of the username in question, this is potentially a serious privacy breach. If he does not know the identity, we show that the attacker may still be able to link the username to an online social network and identify the buyer.

We also show how a large set of eBay buyer usernames can be indirectly obtained from eBay. Given such a large set, an attacker can execute the *broad profiling attack*, namely, determine the purchase history for each of the users in the large set. The attacker can then perform the *category attack*, namely, determine a subset of users who have purchased items in a specific sensitive category, such as gun equipment or medical tests. If the attacker makes the data from the broad profiling attack publicly available, then a third-party can also use side information to de-anonymize a specific target user, giving rise to the *side-information attack*.

In particular, we make the following contributions:

- Show how it is possible to recover a user’s purchase history given his eBay username, despite the privacy measures included in the system.
- Describe several attacks compromising the privacy of eBay users. We discuss three variations: the *broad profiling attack*, the *category attack*, and the *side-information attack*.
- Provide a landscape of user beliefs and expectations regarding eBay privacy, based on a survey of nearly 1,000 subjects.
- Recommend several modifications to the feedback system to allow for better privacy on eBay.

This paper is organized as follows: in Section 2, we introduce the eBay feedback system and some preliminaries. In Section 3, we explain how an attacker can discover the purchase history of a target. In Section 4, we present the broad profiling attack. In Section 5, we describe the category attack, using purchases of gun-related items and medical tests as illustrations. We also briefly discuss the side information attack. Section 6 examines eBay users’ privacy expectations via a survey. In Section 7, we make recommendations to mitigate the risk of privacy attacks. Section 8 summarizes related work. Finally, in Section 9, we conclude.

2 Preliminaries

In this paper, we examine the privacy leaks inherent in eBay’s feedback system. This section describes the eBay feedback system. We also discuss the ethical considerations involved in this research.

2.1 Description of the eBay Feedback System

Feedback Interface The eBay feedback page for a given user is accessible at <http://feedback.ebay.com/ws/eBayISAPI.dll?ViewFeedback2&userid=<username>>, where <username> is replaced with the username in question.

A viewer need not sign in to access a specific user’s feedback page; it is entirely public. As shown in Figure 1, there are several tabs allowing one to filter the feedback shown. One may view all feedback, feedback left on purchases, feedback left on sales, or feedback left for others by the user.

Of particular interest to our work is the tab entitled “Feedback as a Buyer”. This tab displays the feedback left by all sellers from whom the user has made a purchase. Each entry includes the feedback rating (uniformly positive, due to the policies detailed above), the specific feedback message, the seller’s username, and the date and time when the feedback was left. In order to protect the user’s privacy, *no item description or link to the item page is included on the buyer’s feedback page.*

Another tab, entitled “Feedback Left for Others” displays the feedback that the user has left for others. When the user in question is a seller, this primarily contains the feedback he or she has left for customers. Each record includes the item description, a link to the item page, the feedback left, and a pseudonym for the user. *The user’s actual username is not included.*

It is especially important to note that the item’s link can posted even when the buyer does not leave any feedback for the transaction. If the seller leaves feedback (which is estimated to happen in 60-78% of transactions, see Section 8), then the purchase effectively becomes public through no action of the buyer, as we will show.

Public Feedback as a Default As just described, an eBay user’s feedback profile contains a list of the feedback he has given and received. Generally, the comments are public. However, if a user chooses to have a private profile, only his aggregate feedback score is visible; no individual feedback records are shown. eBay states the following regarding feedback profiles³:

Feedback Profiles are public by default. Members have the option of making their Feedback Profiles private. However, it’s important to remember that keeping your profile public builds trust by letting potential trading partners see what others have said about you.


When you choose the “private” setting for your Feedback Profile:

- You can’t sell items on eBay.
- Only the Feedback comments are hidden from other members. Your Feedback Score - the number of positive, neutral, and negative Feedback ratings you’ve received - is still public.

Private Listings Though sellers cannot hide their own feedback history, they can provide additional privacy to buyers by creating a listing with private feedback. Feedback on such a listing will be visible on the seller’s and buyer’s feedback page, but no description or link will be attached to the feedback. Additionally, the bidding history for a private auction is hidden. In all other ways, such as product search and sale procedure, the listing follows standard procedure.

³ <http://pages.ebay.com/help/feedback/profile-public-private.html>

Feedback profile



(4061 ★) me

Positive Feedback (last 12 months): 100%
[How is Feedback percentage calculated?]

Member since: Feb-18-98 in United States

Member quick links

- [Contact member](#)
- [View items for sale](#)
- [View ID history](#)
- [View eBay My World](#)
- [View reviews & guides](#)
- [View About Me page](#)


Feedback as a seller
Feedback as a buyer
All Feedback
Feedback left for others

370 Feedback received (viewing 1-25) **Bid retractions (last 12 months): 0**

Feedback	From Seller	Date/time
+ Super quick payment. Highly-valued eBay customer. Thanks! --	bobslagoon (1474 ★)	Nov-27-13 07:58
+ Thank you for an easy, pleasant transaction. Excellent buyer. A+++++. --	buybest21 (2502 ★)	Aug-22-13 16:46
+ Great repeat buyer - very highly recommended. --	especial_seller (551 ☆)	Jan-04-13 10:44

(a) Buyer Feedback

Feedback profile



bobslagoon (1474 ★) me Top Rated: Seller with highest buyer ratings ?

Positive Feedback (last 12 months): 100%
[How is Feedback percentage calculated?]

Member since: Oct-03-04 in United States

Member quick links

- [Contact member](#)
- [View items for sale](#)
- [View ID history](#)
- [View eBay My World](#)
- [View reviews & guides](#)
- [View About Me page](#)

Feedback as a seller
Feedback as a buyer
All Feedback
Feedback left for others

2,103 Feedback left (viewing 76-100)

Feedback	Left for	Date/time
+ Super quick payment. Highly-valued eBay customer. Thanks! Vintage Tin Heart in Hand Cookie Cutter Nice!! (#190974149951)	Buyer: c***a (4061 ★)	Nov-27-13 07:58 View Item
+ Super quick payment. Highly-valued eBay customer. Thanks! Vintage Advertising Pencil Sharpener "Wayne Feeds" Perfect! (#190974589178)	Buyer: r***e (181 ☆)	Nov-27-13 07:57 View Item
+ Super quick payment. Highly-valued eBay customer. Thanks! Antique Parquetry Box Best!! (#200988161973)	Buyer: o***r (127 ☆)	Nov-24-13 08:50 View Item

(b) Seller Feedback

Fig. 1. Condensed versions of the buyer and seller feedback pages. We have removed the buyer's username and profile picture from the buyer's profile.

Interestingly, eBay advocates limited use of this feature⁴:

While there are some cases where private listings are appropriate, such as the sale of high-priced ticket items or approved pharmaceutical products, you should only make your listing private if you have a specific reason.

Sellers Leaving Feedback for Buyers In the current system, sellers can only leave positive feedback scores for buyers; complaints against buyers are routed through the eBay customer service system instead of being reflected in their feedback. eBay also has additional measures in place to ensure that buyers do not abuse their feedback privileges.⁵

2.2 Ethical Considerations

To implement this research, we built crawlers that visited public eBay feedback pages and downloaded their contents. We then automated content extraction and storage via a customized parser to build inferences from the data.

Performing real-life research in online privacy can be ethically sensitive. Two stakeholders must be considered: the online service provider and the user. While crawling data from online service providers imposes a load upon their servers, we attempted to minimize the load by using a single process to sequentially download pages. Regarding the user, we point out that any inferences we made were based on publicly available data; however, we have taken steps to store our data in a secure manner.

Moreover, this research benefits the eBay ecosystem by encouraging more private methods of displaying feedback. Users benefit from increased privacy measures, and eBay may benefit since users are more likely to buy from online retailers who visibly promote privacy, as shown by Tsai et al. [28].

3 Recovering Purchase History

In this section, we detail the purchase history attack, namely, how an attacker can recover the purchase history of a target when given the target's username.

At first glance, it does not seem possible to recover a user's purchase history from the feedback pages. Indeed, on the buyer's page, the items that the buyer bought are not listed; on the seller's page, although the items sold are listed, the buyers of the items are not provided. However, we show that a buyer's purchase history can be determined by exploiting the timestamp information on the feedback pages.

Each feedback record is displayed with a timestamp, both on the seller's page and the buyer's page. This allows for linking of feedback records from a seller's account to a buyer's account through the following process:

⁴ <http://pages.ebay.com/help/sell/private.html>

⁵ <http://pages.ebay.com/services/forum/sellerprotection.html>

1. Retrieve the user’s feedback page.
2. Extract the seller’s name and the timestamp for each feedback entry.
3. For each feedback entry, visit the seller’s page. Then search among the feedback listings for feedback with an identical timestamp. Retrieve the item link and description.
4. Output the list of the user’s sale records.

However, in some cases a seller may have left feedback for more than one purchase simultaneously (perhaps through an automated system). Thus, relying solely on the timestamp may introduce false records into the target’s purchase history. To study this issue, we examined 5,580 randomly chosen purchases. We found that 49% of the timestamps on buyers’ pages matched with only one distinct listing from the seller’s feedback page. On average, each buyer feedback record matched the timestamps of 6.5 records from the seller’s feedback; the median was 2 matches. In one specific case, the timestamp on one buyer’s corresponded to as many as 279 feedback records from the seller in question. (The buyer in this case had made several purchases from a seller who used an automated system to post large batches of feedback.) To resolve this ambiguity that occurs in approximately half of the transactions, we extend the above attack by leveraging the pseudonyms included in the seller’s feedback page.

While the seller’s feedback page uses only a pseudonym to identify the buyers, each user’s pseudonym remains consistent across the site. eBay assigns pseudonyms according to a specific algorithm: randomly select two character’s from the user’s real username and insert three asterisks in between them to form the pseudonym⁶. This allows an attacker to definitively rule out any pseudonyms that could not be generated by a specific username. For example, if the targeted user goes by the user ID “catlady24”, then the pseudonym “u***v” cannot correspond to that user.

The number of possible pseudonyms per username is bounded by $n(n - 1)$, where n is the length of the username. As such, the pseudonym is not random, but is rather chosen from a relatively small space of potential pseudonyms.

Based on this additional data, we modify the above process for purchase recovery for a given user to reduce false associations:

1. Retrieve the user’s feedback page.
2. Extract the seller’s name and the timestamp for each feedback entry.
3. For each feedback entry, visit the seller’s page. Then search among the feedback listings for feedback with an identical timestamp. Retrieve the item link and description.
4. When all the purchases are retrieved, remove all feedback entries which have pseudonyms that could not be generated by the username.

By utilizing the pseudonym as a heuristic to rule out listings with invalid pseudonyms, we were able to reduce the number of potential matches in our

⁶ <http://community.ebay.com/t5/Bidding/Bid-History-Changes-including-a-b-userIDs/m-p/2443087#M26865>

sample database by roughly 70%. However, after filtering by timestamp and invalid pseudonyms, there were still false matches remaining in the database, with an average of 1.9 potential matches for each listing in a buyer’s feedback. To reduce the number of false matches, we leverage the fact that a user’s pseudonym is consistent across the feedback system. Since each user has only one actual pseudonym in the system, we attempt to find this pseudonym and thus eliminate any potential matches using other pseudonyms. In our sample database, 73% of users had more than one potential pseudonym remaining at this point in the process. We aim to resolve this ambiguity with the following steps:

5. If more than one pseudonym remains among the buyer’s matched records:
 - (a) Conduct a vote where each seller nominates the pseudonym that dominates its corresponding records for the user.
 - (b) Select as the correct pseudonym the one which has the most votes.
 - (c) Eliminate all records which use a different pseudonym
6. Output the list of the user’s sale records.

Through the steps above, it is possible to recover both a user’s purchase history and their pseudonym, given their real username. Not only does this allow one to see the user’s past purchase behavior, it makes it easier to monitor future behavior since the attacker has learned the user’s persistent pseudonym.

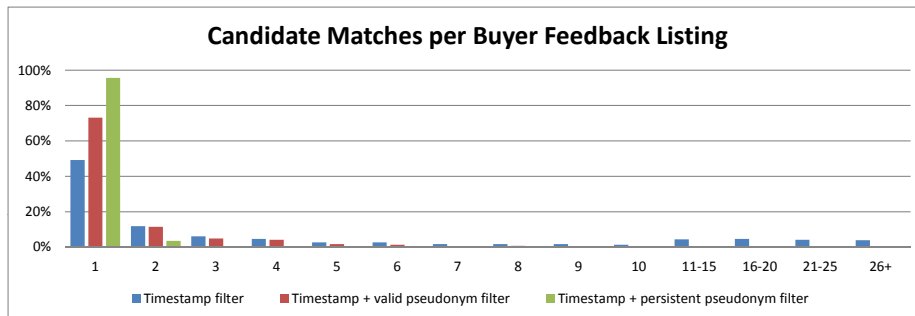


Fig. 2. The distribution of matches found per buyer feedback listing when using the different filtering methods. The precision of the matches increases considerably with the more advanced filtering methods.

When testing the database of 5,580 feedback records, extending the technique with pseudonym information enabled us to match 96% of buyers’ feedback records to a single seller feedback record complete with purchase details. Likewise, we were able to learn a single persistent pseudonym for 96% of the sampled users. Figure 3 shows the how the modifications to the filtering method reduce the number of matches found per purchase.

4 Broad Purchase Profiling

To further illustrate the privacy leakage potential of the eBay feedback system, we introduce a broad purchase profiling attack where we find users on eBay and associate their purchase history with a real name drawn from Facebook.

4.1 Motivation

The ability to collect widespread eBay purchase data and associate it to real people is of use to several actors. Advertisers and content publishers would like to collect user purchasing behavior in order to present targeted ads, and marketers would like to analyze which purchases are bought together in order to aim their products at specific segments. Additionally, companies providing background checks for employers or insurance companies may want to include purchasing behavior in their classification methods. Finally, malicious parties may want to build detailed dossiers on eBay users in order to enable sophisticated spear phishing attempts.

In each of these cases, eBay feedback information can be utilized to engineer a privacy breach by inferring potentially sensitive facts about users which were not previously known. The association of these records to real people constitutes a privacy liability.

4.2 Execution

Given a list of eBay user IDs, we detail how to infer the name and purchase history of each user. Here, we make the assumption that the attacker has access to a substantial amount of computing and bandwidth resources. Also, when crawling eBay, we assume the attacker is clever enough to introduce sufficient delays between queries so that eBay does not block his requests; to expedite the attack, he may also use multiple IP addresses.

The first step of the attack involves identification of users' real names. To accomplish this, we leveraged the Facebook Graph API⁷, a tool for building applications integrated with the Facebook social graph. (Using the Graph API via a browser does not require a developer account; however, integrating it into an automated crawler program requires developer and app tokens, which can be accessed for free after a short sign-up procedure requiring only a Facebook account.)

To test each eBay username for a match, we sent an HTTP request to `http://graph.facebook.com/<username>`, where `<username>` was replaced with the eBay username in question. If a match was found, then a response (pictured in Figure 3) was received, detailing the matched account's name, gender, locale, a unique numerical Facebook ID, and (in most cases) a link to their profile.

⁷ <http://developers.facebook.com/docs/reference/api/>

This produced a list of eBay user IDs and corresponding potential real names. To recover the users' feedback history, we followed the process in Section 3, but with a slight modification to reduce the network and processing overhead. Instead of individually retrieving the sellers on a buyer-by-buyer basis, and thus perhaps retrieving duplicate pages, each buyer page was parsed to compile a comprehensive list of all sellers who had engaged with our list of buyers. We then downloaded each seller's page once and proceeded with the above methods of matching purchase histories to users.



Fig. 3. A sample response from the Facebook Graph tool. Note that the tool performs some simple string-matching to match the term “johnsmith” with the account “John.Smith”.

4.3 Results

We began with a list of 130,991 usernames. In order to limit the extent of our crawling, we first attempted to match the usernames to Facebook accounts and names before proceeding with purchase profiling. 22,478 matches were found, for a match rate of 17.2% (see Table 1). It is important to note that this method does not conclusively match accounts. For example, the accounts using a common name such as “bob123” on Facebook and eBay may very well belong to two different people [22]. As such, the match rate of 17.2% should be considered an upper bound which includes some false positives.

The 22,478 usernames were then used to recover their recent purchase history. In total, this produced a list of 414,483 purchases, for an average of approximately 18 purchases per customer. Matching the information from user feedback with seller feedback provided item descriptions and links, based on the timestamp and pseudonym information.

In summary, we were able to match 17% of 130,991 users with a potential real name on Facebook. For the matched users, we discovered on average 18 purchases per person, complete with purchase description.

Case Study	List of User IDs	Names Found	Match Rate
Broad user profiling	130,991	22,478	17.2%
Gun accessories	228,332	35,262	15.4%
Pregnancy tests	27,261	4,694	17.2%
H.I.V. tests	221	37	16.7%

Table 1. The number of user IDs, Facebook names matched, and the match rate for each experiment.

4.4 Collecting Usernames

The above method assumes that the attacker has access to a large set of eBay usernames. An attacker can obtain a list of eBay usernames through several avenues:

- **eBay social features:** eBay provides a feature where buyers can follow sellers. Each seller’s list of followers is publicly accessible at <http://www.ebay.com/usr/<sellername>/followers>, where <sellername> is replaced with the seller’s user ID. The buyer IDs are provided in cleartext. This enables an attacker to crawl buyers’ user IDs by iteratively visiting a seller page to collect followers and then collecting more sellers from each of the follower’s feedback pages.
- **Seller’s names:** eBay users can both buy and sell; many eBay users engage in both activities. It is possible to discover eBay sellers via the search interface. By searching for many items, one can thus compile an extensive list of eBay sellers and then determine their purchase histories. As we show via a survey in Section 6, 60% of eBay users have sold at least one item; therefore, this approach can expose many usernames.
- **Username reuse:** the attacker can use a list of usernames that are publicly available on some other service (such as Twitter or Facebook⁸). By using the feedback interface as an oracle, the attacker can determine if the usernames are in use at eBay as well.
- **Social engineering:** the attacker can sell popular items on eBay to collect eBay usernames (since sellers can view the usernames of users who purchased from them).
- **Brute force:** the attacker can generate potential usernames and use the feedback interface as an oracle to determine if the usernames are in use.

5 Category and Side Information Attack

Beyond exposing the feedback history of many users, the feedback history can also be exploited to generate a list of users who have engaged in specific, potentially sensitive categories of transactions. In this section, we elaborate an attack

⁸ <http://www.facebook.com/directory/>

to reveal users who have bought specific items from eBay. This more vividly demonstrates the privacy risks embodied in the eBay feedback system.

This method can be accomplished using the basic techniques laid out in Section 4. Here, we introduce some modifications to streamline the attack further. Using the eBay search interface, it is possible to perform keyword search for completed sales of a specific item. This provides a list of sellers who sell the merchandise in question. This list allows the attacker to reduce the network overhead of his attack; instead of attempting to discover the purchase history of all the user IDs in his database, he can restrict his search to the users who have interacted with the sellers who sell this item. The process would be as follows:

1. Using eBay’s keyword search, curate a list of sellers who have sold the item in question.
2. For each buyer in the database, download their feedback page.
3. For each buyer’s feedback, discard any feedback not related to a seller on the list.
4. For the remaining records, recover the purchase history as detailed in Section 3.

We now introduce a few case studies that are enabled by this approach.

5.1 eBay Gun Registry

Gun control and ownership in the United States is a highly charged topic, as illustrated by the public uproar in December 2012 when a newspaper in Westchester County, NY published a list of local gunowners and their addresses [30]. Readers and residents considered this to be a massive privacy breach and the paper later removed the list [11].

Crawling eBay to find gun owners could be of interest to several parties. Firstly, law enforcement or private investigators may want to search for unregistered gun owners. Secondly, background check providers or data aggregators may want to include gun ownership in their records. In certain municipalities, a list of registered gun owners in a county can be accessed by filing a Freedom of Information request; however, the approach described here has several advantages. Firstly, it leaves no legal trail, a fact which may be appealing to actors (both criminal and innocent) operating outside of the legal framework. Secondly, it can help to uncover unregistered gun owners who have purchased gun supplies or accessories online. Thirdly, since guns are generally registered at the local level, compiling an extensive list would require many FOIA requests to different authorities; this approach is not subject to such limitations.

We began our data collection of eBay users by searching eBay for purveyors of gun accessories. (Actual firearms are not sold on eBay, so we use purchase of firearm accessories, such as holsters, as a proxy for gun ownership.) A search of “gun holster” on eBay’s web site enabled us to curate a list of 11 sellers who sell firearm accessories at high volume.

Using a second external list of usernames, we narrowed down the users to those who had interacted with the identified gun-accessory sellers. For each

matched purchase, a simple text-based classifier was applied to ascertain that the purchase was gun-related. Afterwards, we used Facebook’s Graph tool to find potential names for each buyer.

Based on a list of 11 firearms-accessory sellers, we found records of 292,827 gun accessory purchases, made by 228,332 unique buyers. After matching for names on Facebook, we had names and sale records for 35,262 suspected gun owners.

5.2 Medical Test Purchases

In this second case study, we present an attack against users who have purchased sensitive medical tests, namely pregnancy or fertility tests and H.I.V. tests.

As a health-related issue, pregnancy is a private concern for many people. Pregnant women may wish to keep their reproductive status secret for a variety of reasons: they may be concerned about the pregnancy’s medical viability, afraid of job discrimination or social stigma, or desirous of quietly ending the pregnancy without censure. In nearly all cases, expectant parents prefer to have control over who gains access to their news. This issue was highlighted when the retail giant Target began to track buyers’ purchases in order to predict pregnancy among buyers (and subsequently advertise to them); buyers were unsettled by the fact that a retail store had estimated their due date even before they had told their parents [8]. Pregnancy is clearly private information, and people who are pregnant or trying to conceive deserve privacy for related purchases.

Even more private is H.I.V.-positive status. People with H.I.V. may find themselves subject to discrimination in a number of ways due to social stigma related both to the disease itself and its associated risk factors. As such, it is of prime importance that purchases related to H.I.V. testing remain hidden from the public eye.

These attacks were executed in the same manner as the gun registry attack. Specifically, we collected a list of eBay accounts that sell tests for these medical conditions and then matched them against a list of users’ feedback pages to find users who had bought these items. Having found eBay usernames, we then fed these identifiers to the Graph API to find matching Facebook accounts and real names.

After building a list of pregnancy test sellers, comparing with a list of usernames and feedback, and filtering out non-fertility-related purchases, we had collected a list of 27,261 unique eBay users that purchased fertility-related tests. (We use this term to include ovulation tests, pregnancy tests, and gender-prediction tests.) Of these users, we found 4,694 matching Facebook accounts for a matching rate of 17.2%. This supplied us with a list of nearly 5,000 real names for potentially pregnant people.

The number of users returned for the targeted pregnancy test crawl was much smaller than the number returned for the gun crawl. This can be explained by the observation that, as of the the time of writing, there are more than 15 as many listings on eBay for the search “gun holster” than for the search term

“pregnancy test.” This indicates a greater market for gun accessories than for medical tests on eBay, enabling a larger-scale crawl for gun-related purchases.

The crawl for buyers of H.I.V. test yielded much more meager results, for a number of reasons. Firstly, H.I.V. is a much rarer condition than pregnancy, and therefore is not as commonly tested. Secondly, H.I.V. tests are significantly more expensive than pregnancy tests and therefore less likely to be bought often. Thirdly, H.I.V. tests are commonly administered in a medical setting, not at home (unlike pregnancy tests). Fourthly, many sellers of H.I.V. tests had taken precautions to keep the listings private.

Nonetheless, an extensive crawl did discover 221 unique users who had purchased H.I.V. tests online. 37 of these users were matched to accounts and names on Facebook. Of the matched users with a specified gender, 28% were female and 72% were male; this correlates roughly with the general rates of H.I.V. infection in the USA [4].

5.3 Side Information Attack

We now briefly describe another possible attack, utilizing side information. For concreteness, we state this problem in terms of employer who is suspicious of his employee with regard to health, interests, or other sensitive information that could be revealed in eBay purchase data. The employer knows that his employee uses eBay, but he does not know his employee’s eBay user name. Finally, suppose he has some side information; namely, that his employee made a specific purchase on a specific day.

Now assume that an attacker has carried out the broad profiling attack on a set of users containing the employee and makes the corresponding database publicly available on a website.

For each piece of side-channel purchase data that the boss knows about his employee, he can narrow down the set of prospective matches. Intuitively, the more side information he has, the smaller the set of candidate matches will become. Once he has a small enough set of candidate matches, he can then attempt to pinpoint a specific username as corresponding to his target. This enables him to learn the entire purchase history of his target, and it also makes it easy for him to continue monitoring any future eBay purchases made by the employee.

6 User Expectations for Privacy on eBay

Are users aware of the eBay feedback policies, and how accessible do they expect their purchase data to be? We conducted a survey of eBay users on Amazon Mechanical Turk⁹, a crowdsourcing microtask market, to answer these questions. To maintain a uniform high quality of responses, we followed the guidelines established by Kelley and Patrick [14] while designing our survey to make sure that users stayed engaged, attentive, and honest.

⁹ <http://www.mturk.com>

6.1 Survey Design

Our questions were designed to answer a few specific questions about eBay users’ expectation and behavior.

- Are usernames considered private information by users?
- Is eBay a place where people purchase sensitive items?
- Whom do users expect to be able to see their purchase history?

We limited participation in our study to US-based Mechanical Turk workers. We screened subjects to make sure they were actually users of eBay. We also incorporated demographic questions. Each worker was paid \$0.25 for a multiple-choice survey that took 3 or 4 minutes to complete.

6.2 Survey Results

We gathered 1114 responses from Mechanical Turk. After removing “click spam” as detected by attention-measuring questions, we had 913 responses. To assess the representative qualities of our sample, we compared the reported demographics to those measured by the Google AdWords Display Planner Tool¹⁰. The gender proportions of our sample exactly matched the eBay users measured by AdWords, and the sample also followed the general age trends reported by AdWords, albeit with less precision. Some variance may be due to the generally younger population of Mechanical Turk workers, as measured by Ross [24]. Overall, the sample proved to be highly representative of the general eBay population.

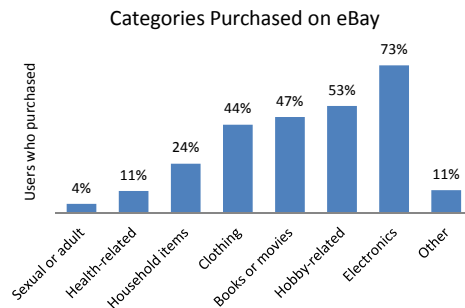


Fig. 4. The percentage of users who have bought items from specific categories on eBay.

Purchase Behavior on eBay Do users buy sensitive items on eBay? We asked our subjects if they had purchased items from several categories, listed in Figure

¹⁰ <https://adwords.google.com/da/DisplayPlanner/Home>

4. Notably, 10.7% of users answered that they had bought health-related items on eBay, and 4.3% of users had bought sexual/adult items on eBay.

While we consider health- and adult-related purchases to be uniformly private information, there may also be sensitive purchases made in other categories. In order to gain a broader view of how users view eBay, we asked them to answer the following question: “If I need to buy something sensitive or embarrassing, I would probably buy it from...” The available answers were “a physical store,” “eBay,” and “other (please specify).” eBay edged out physical stores as the most likely place to buy sensitive items. 38.6% of users chose eBay, 35.6% of users chose a physical store, and 25.8% chose other. From these numbers, it is clear to see that users see eBay as a leading way to easily and anonymously conduct transactions that may be embarrassing or sensitive.

Users and Feedback Do eBay users leave feedback? Do they understand the way it works? Our survey included a series of questions about usage and perception of the feedback system. 50.7% of the respondents reported that they leave feedback all the time, with only 10.7% answering that they never leave feedback. The remaining respondents have left feedback somewhat often (17.4%) or a few times (21.1%). Despite their extensive use of the feedback system, users did not seem to understand the complexities of the system. Only 18% of users correctly agreed with the statement that “sellers cannot leave negative feedback for buyers on eBay”. 66.4% disagreed with the statement, and 15.7% were not sure. This is despite the fact that 63.3% of the users reported that they had sold at least one item on eBay.

Privacy Expectations Users understood that feedback was public: 76.6% of users agreed with the statement that “anyone can see the feedback that sellers leave on my account.” However, it was unclear to many users just how this related to their purchase privacy. In answer to the question, “Who can see the purchases I make on eBay?” the largest portion (38%) of users selected “just me.” Only 8.7% of users selected the most correct answer, “anyone, even if they are not signed in.” See Figure 5 for a breakdown of their answers.



Fig. 5. Users selected one answer that they believed to best include who could view their purchases on eBay. The correct answer is highlighted.

7 Recommendations

In many ways, eBay’s system for buyer reputation seems to be an artifact of an older era. Since 2008, the feedback system on eBay allows sellers to leave only positive feedback for buyers. As such, a buyer’s feedback score reflects merely the quantity, not the quality, of his transactions.

Additionally, buyers’ reputations matter less to the success of the eBay marketplace than those of sellers. In contrast to sellers, who can easily defraud consumers by failing to ship or shipping incorrect items, buyers have a much more limited scope of possible fraud. While a buyer might fail to pay, eBay policy advises sellers not to ship their wares until buyers have paid. Additionally, there are internal methods by which sellers can report fraudulent buyers.

Allowing sellers to leave feedback about buyers may benefit the eBay marketplace by encouraging buyers to then reciprocate with feedback about the buyers. Moreover, since feedback from purchases counts towards the user’s general feedback score, it may give experienced eBay shoppers an advantage when they first begin to sell goods on eBay. However, it is unclear why feedback left by sellers for buyers should be made public. Since the feedback is uniformly positive, it does not offer helpful guidance for future transactions. Considering the privacy leakage which public buyer reputations enable, we make the following recommendations:

- **Recommendations to eBay:** We recommend that the Private Listing option should become the default listing method. This removes any visible link between the buyers and sellers. Alternatively, we propose that eBay use a non-persistent pseudonym for buyers on the sellers’ feedback pages; this would make it harder to link feedback. Additionally, the timestamp of the feedback left could be generalized (for example, by displaying only the date) in order to make linkage attacks more difficult.
- **Recommendations to buyers:** Buyers on eBay can make their feedback profile private by changing a setting in their Feedback Forums page¹¹. However, accounts with private feedback profiles cannot be used for selling. We therefore recommend that eBay users maintain two separate accounts, a private profile for buying and a public account for selling. (However, this does not obviate the need for an eBay policy change, since it prevents a user’s selling account from reaping the benefits of the positive feedback he has earned as a buyer on eBay.)

We also recommend that users avoid reusing usernames across different websites in order to retain stronger pseudonymity.

- **Recommendations to sellers:** eBay offers a selling option called a Private Listing¹² which operates exactly like a regular listing while keeping all buyer information anonymous. This is a way for sellers to offer their buyers all buying benefits while retaining privacy. Other users cannot see the list of bids on the item, and all feedback on the item is anonymous for feedback

¹¹ <http://pages.ebay.com/help/feedback/profile-public-private.html>

¹² <http://pages.ebay.com/help/sell/private.html>

profile views. The listing is included in searches as usual, and it has no extra fees associated with it.

8 Related Work

Feedback and Reputational Systems The eBay reputation system has been studied in depth by the economics and management science communities to assess the impact of ratings on sale statistics. Lucking-Reiley et al. [17] found that increases in negative ratings were associated with significantly lower selling prices. While Resnick et al. [23] found that the impact of reputation was inconsistent across areas, Houser et al. [12] found that better seller reputations tended to raise prices in auctions.

How widely is the eBay feedback system used? Do buyers and sellers take the time to submit ratings and comments? Based on a 1999 dataset of more than 30,000 postcard sales, Resnick et al. [23] found that 52.1% of buyers and 60.8% of sellers left feedback on transactions. In 2008, Dellarocas and Wood [7] found in a study of over 50,000 rare coins that 78% of sellers and 68% of buyers left feedback on a purchase. 20% of the auctions in their dataset had feedback from sellers only, with no feedback from buyers (as opposed to feedback from both, neither, or only the buyer).

Klein [15] found that buyers' fear of retaliatory feedback from sellers led them to leave feedback at the last available moment, and may have even suppressed feedback rates due to fear of retaliation. Chwelos and Dehar [5] found that the two-way nature of the system, where sellers and buyers rate one another, has a dual effect: it encourages buyers to leave more feedback, but it also inflates positive feedback. Klein et al. [16] suggest that removing the ratings of buyers would create a less sugar-coated feedback system while introducing little risk, since there is little need for the seller to trust buyers in any case.

Deanonymization Robust data anonymization has been a high-profile problem since 2006, when a woman was identified based on her "de-identified" queries in the publicly released AOL search logs [3]. External data sources were also used by Narayanan and Shmatikov [19] to reidentify the anonymous NetFlix dataset with matching members on IMDB, using only the sparse vector of movies watched and reviewed. Approaches for deanonymization of social networks are presented by Backstrom et al. [2], Narayanan and Shmatikov [20], and Wondracek et al. [29]. Goga et al. [10] correlate user accounts from Twitter, Flickr, and Yelp by using only temporal, geographic, and language features.

These attacks have given rise to several frameworks with the aim of providing a more disciplined and guarantee-based approach to data anonymization. These approaches add anonymity at the expense of specificity; namely, they generalize the dataset until there are fewer distinctive records. The first of these approaches, by Sweeney, was k -anonymity [27], which iteratively generalizes information until there are at least k records present in the dataset that match any specific tuple. More recently, Machanavajjhala et al. [18] proposed l -diversity, which

extends k -anonymity to disallow uniform generalizations about specific populations. Finally, differential privacy, by Dwork [9], advocates systematically adding Laplacian noise to query answers in order to provide numeric privacy guarantees that can quantify the risk of a privacy breach.

Username Reuse While users are warned about the dangers of password reuse across online accounts (see for example Ives et al. [13]), there is little work discussing the phenomenon of username reuse. Perito et al. [22] research the entropy of usernames, finding that certain usernames are more unique (and thus more traceable) than others. Using measures such as Levenshtein Distance and TF/IDF, they were able to achieve a recall of 71% in matching related usernames.

Privacy-Preserving Reputation Systems For the most part, reputation systems have focused on accountability rather than privacy. However, there are cases where reviewers may wish to hide their identity, both from other users and from a centralized authority. This has led to the research topic of decentralized reputation systems [26]. Schemes have been implemented to retain properties such as security against forged reviews and persistent reputations in the face of multiple pseudonyms to allow for anonymity [1].

Notably, Pavlov et al. [21] propose a decentralized system allowing for privacy on the reviewer’s part as well as easy additive aggregation of users’ reputations from across the decentralized system. This deals with the problem of privacy from the reviewer’s perspective; however, it does not focus on privacy breaches from the viewpoint of the party receiving feedback.

Schiffner et al. [25] point out theoretical limits on the utility of any fully private reputational system. Consequently, in [6], Clauß et al. construct a reputation system that conforms to the more relaxed privacy definition of k -anonymity.

Our Contributions Existing research has explored the utility of reputation systems; however, there has not been any in-depth investigation into the privacy implications of the eBay feedback system. What risks are inherent in its setup, and how can they be exploited? How does this comply with user expectations of privacy? In this work, we explore these questions and offer recommendations based on our findings.

9 Conclusion

This research brings to light several important issues. Firstly, we show how an attacker can determine a target’s purchases if he knows the target’s eBay username. Though the feedback interface does not explicitly link buyers to purchases, we leverage feedback timestamps and pseudonym information to infer a list of purchases made by a user. We present several classes of attacks complete with case studies to show how serious the breach is. In the feedback history attack, we show how to recover one user’s history. In the broad profiling attack, we modify the technique to recover purchase histories of many users. In the category

attack, we show how to collect a list of users who have bought specific items. We demonstrate this attack by uncovering buyers of gun accessories, pregnancy tests, and H.I.V. tests. Subsequently, we examine user expectations of privacy on eBay and find a serious clash with reality; eBay is much less private than users believe. Finally, we recommend several techniques to mitigate the privacy risks of the system.

Acknowledgements

This work was supported in part by the NSF under grants CNS-1318659 and 0966187. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of any of the sponsors.

References

1. E. Androulaki, S. G. Choi, S. M. Bellovin, and T. Malkin. Reputation systems for anonymous networks. In *Privacy Enhancing Technologies*, pages 202–218. Springer, 2008.
2. L. Backstrom, C. Dwork, and J. Kleinberg. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of the 16th international conference on World Wide Web*, pages 181–190. ACM, 2007.
3. M. Barbaro, T. Zeller, and S. Hansell. A face is exposed for aol searcher no. 4417749. *New York Times*, Aug. 9, 2006.
4. CDC. H.i.v. incidence, May 22, 2013.
5. P. Chwelos and T. Dhar. Caveat emptor: Differences in online reputation mechanisms. Technical report, Working Paper, Sauder School of Business, University of British Columbia, 2006.
6. S. Clauß, S. Schiffner, and F. Kerschbaum. k-anonymous reputation. In *Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security*, pages 359–368. ACM, 2013.
7. C. Dellarocas and C. A. Wood. The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science*, 54(3):460–476, 2008.
8. C. Duhigg. How companies learn your secrets. *New York Times*, Feb. 16, 2012.
9. C. Dwork. Differential privacy. In *Automata, languages and programming*, pages 1–12. Springer, 2006.
10. O. Goga, H. Lei, S. H. K. Parthasarathi, G. Friedland, R. Sommer, and R. Teixeira. Exploiting innocuous activity for correlating users across sites. In *Proceedings of the 22nd international conference on World Wide Web*, pages 447–458. International World Wide Web Conferences Steering Committee, 2013.
11. J. D. Goodman. Newspaper takes down map of gun permit holders. *The New York Times*, Jan. 18, 2013.
12. D. Houser and J. Wooders. Reputation in auctions: Theory, and evidence from ebay. *Journal of Economics & Management Strategy*, 15(2):353–369, 2006.
13. B. Ives, K. R. Walsh, and H. Schneider. The domino effect of password reuse. *Communications of the ACM*, 47(4):75–78, 2004.

14. P. G. Kelley. Conducting usable privacy & security studies with amazons mechanical turk. In *Symposium on Usable Privacy and Security (SOUPS)*, 2010.
15. T. Klein, C. Lambertz, G. Spagnolo, and K. O. Stahl. Last minute feedback. *Centre for Economic Policy Research*, 2006.
16. T. J. Klein, C. Lambertz, G. Spagnolo, and K. O. Stahl. The actual structure of ebay’s feedback mechanism and early evidence on the effects of recent changes. *International Journal of Electronic Business*, 7(3):301–320, 2009.
17. D. Lucking-Reiley, D. Bryan, N. Prasad, and D. Reeves. Pennies from ebay: The determinants of price in online auctions*. *The Journal of Industrial Economics*, 55(2):223–233, 2007.
18. A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3, 2007.
19. A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *IEEE Symposium on Security and Privacy, 2008*, pages 111–125. IEEE, 2008.
20. A. Narayanan and V. Shmatikov. De-anonymizing social networks. In *IEEE Symposium on Security and Privacy, 2009*, pages 173–187. IEEE, 2009.
21. E. Pavlov, J. S. Rosenschein, and Z. Topol. Supporting privacy in decentralized additive reputation systems. In *Trust Management*, pages 108–119. Springer, 2004.
22. D. Perito, C. Castelluccia, M. A. Kaafar, and P. Manils. How unique and traceable are usernames? In *Privacy Enhancing Technologies*, pages 1–17. Springer, 2011.
23. P. Resnick and R. Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of ebay’s reputation system. *Advances in applied microeconomics*, 11:127–157, 2002.
24. J. Ross, L. Irani, M. Silberman, A. Zaldivar, and B. Tomlinson. Who are the crowdworkers?: shifting demographics in mechanical turk. In *CHI’10 Extended Abstracts on Human Factors in Computing Systems*, pages 2863–2872. ACM, 2010.
25. S. Schiffner, A. Pashalidis, and E. Tischhauser. On the limits of privacy in reputation systems. In *Proceedings of the 10th annual ACM workshop on Privacy in the electronic society*, pages 33–42. ACM, 2011.
26. M. Srivatsa, L. Xiong, and L. Liu. Trustguard: countering vulnerabilities in reputation management for decentralized overlay networks. In *Proceedings of the 14th international conference on World Wide Web*, pages 422–431. ACM, 2005.
27. L. Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.
28. J. Y. Tsai, S. Egelman, L. Cranor, and A. Acquisti. The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, 22(2):254–268, 2011.
29. G. Wondracek, T. Holz, E. Kirda, and C. Kruegel. A practical attack to de-anonymize social network users. In *IEEE Symposium on Security and Privacy, 2010*, pages 223–238. IEEE, 2010.
30. D. R. Worley. The gun owner next door: What you don’t know about the weapons in your neighborhood. *The Journal News*, Dec. 24, 2012.