Privacy Threats in E-Shopping (Position Paper)

Jesus Diaz[1], Seung Geol Choi[2], David Arroyo[1], Angelos D. Keromytis[3], Francisco B. Rodriguez[1], and Moti Yung[3,4]

[1] Universidad Autónoma de Madrid
{j.diaz,david.arroyo,f.rodriguez}@uam.es
[2] United States Naval Academy, choi@usna.edu
[3] Columbia University, {angelos,moti}@cs.columbia.edu,
[4] Google Inc.

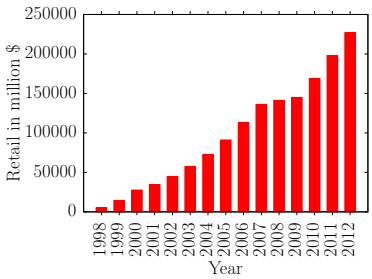
Abstract. E-shopping has grown considerably in the last years, providing customers with convenience, merchants with increased sales, and financial entities with an additional source of income. However, it may also be the source of serious threats to privacy. In this paper, we review the e-shopping process, discussing attacks or threats that have been analyzed in the literature for each of its stages. By showing that there exist threats to privacy in each of them, we argue our following position: "It is not enough to protect a single independent stage, as is usually done in privacy respectful proposals in this context. Rather, a complete solution is necessary spanning the overall process, dealing also with the required interconnections between stages." Our overview also reflects the diverse types of information that e-shopping manages, and the benefits (e.g., such as loyalty programs and fraud prevention) that system providers extract from them. This also endorses the need for solutions that, while privacy preserving, do not limit or remove these benefits, if we want prevent all the participating entities from rejecting it.

Keywords: Privacy, Online shopping, Payment systems, Purchase systems.

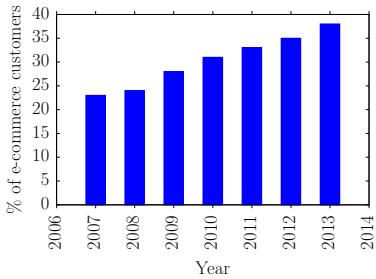
1 Introduction

E-shopping[5] has been growing continuously (Figure 1), providing customers with convenience, merchants with increased sales, and financial entities with an additional source of income. Concurrently, e-shopping has become more complex, the complete process being divided in several stages dealing with specific subprocesses (i.e., purchase, payment, delivery and completion). As the e-shopping process has grown more complex, protecting privacy of consumers has become more difficult. There are multiple parties involved, each managing various pieces of information; if any single party mistreats or misuses consumer data, consumer privacy will be at risk. Moreover, the complexity of the information each party manages makes the mistreatment or misuse only more likely.

[5] In this work, we restrict ourselves to the context of B2C (business-to-consumer). B2B (business-to-business) may require additional considerations.



E-commerce retail sales in USA between 1998 and 2012 [19].



% of EU-28 users having bought via e-commerce in last 3 months [10].

Fig. 1: Indicators of e-commerce growth in USA and EU-28.

In this paper, we review the e-shopping process, discussing attacks or threats analyzed in the literature for each of its stages. By showing privacy threats in each phase, we argue that it is not enough to protect a single independent stage, as is usually done in privacy respectful proposals in this context. Rather, a complete solution is needed spanning the overall process, dealing also with the required interconnections between stages. Our overview also reflects the diverse types of information that e-shopping manages, and the benefits (like loyalty programs and fraud prevention) that system providers extract from them. This endorses the need for solutions that, while privacy preserving, do not hinder these benefits, if we want prevent all the participating entities from rejecting it.

2 The Process of E-Shopping Transactions

The participants in e-shopping are basically the same as in the conventional shopping setting. Customers (`C` hereafter) acquiring goods, merchants (`M`) offering their products, and banks, credit card companies, etc., responsible for managing the financial backend (hereafter referred to as financial network, or `FN`). Finally, when selling physical goods, a delivery company (`DC`) is also necessary.

The entire process of an e-shopping transaction may be divided in three phases [24], plus an optional final phase (see Figure 2):

1. Purchase. Customer `C` selects the products from merchant's `M` website.
2. Checkout. Having specified the shipping and payment information, `C` confirms the purchase and pays (through `FN`) for the selected products.
3. Delivery. `M` (probably, through `DC`) delivers the products to `C`.
4. [Optional] Completion (evaluation and dispute solving). `C` evaluates her experience, maybe including product returns or refunds.

3 Threats to Privacy in E-Shopping

We review existing threats to privacy through the lens of e-shopping as described in Section 2. Table 1 summarizes the threats that have been exposed to some extent in the literature for each phase (in the references under the third column).

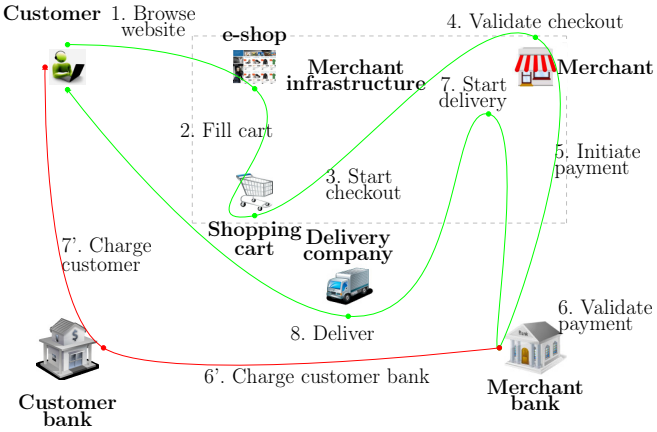


Fig. 2: Overview of the online shopping process.

We note that some of them occur quite naturally in current industry systems (like threat 2.2, since M usually learns C's payment information, or threat 3.2, since DC learns both C and M addresses, being able to link C and M). However, as we will see in some of the reviewed attacks, sometimes it is enough with few additional information in order for an attacker to seriously undermine privacy. Therefore, it is advisable to keep the principle of *least information*. We also note that we mostly focus on threats which allow third (or unauthorized) parties to gain information they should not learn in a privacy-ideal scenario; many times, this leads to threats to anonymity, or leakage of product or payment information.

Threats in the purchase stage. We first note that 9 out of 13 risks outlined in [5] (risks 4 to 13, mostly dealing with C's personal information and M's dishonesty) affect the purchase process and the communications between C and M.

Additional vulnerabilities may lie in loyalty programs; M can apply loyalty programs to lure more customers into buying products. The promotions offered by M will probably be based on C's profile, purchase history, and shopping cart. Since M will likely apply loyalty programs to increase their revenue, it is necessary

Phase	Privacy threats	References
Purchase	1.1. Product info leaked to FN or 3rd parties 1.2. Link C and M by 3rd parties	[5,22]
Checkout	2.1. Product info leaked to FN or 3rd parties 2.2. Payment info leaked to M or 3rd parties 2.3. Link C and M by 3rd parties	[25,16,5,1,2,8,15,22]
Delivery	3.1. Shipping address leaked to M or 3rd parties 3.2. Link C and M by DC or 3rd parties	[4,5]
Completion	4.1. Private info leaks through feedback (All previous threats may affect completion)	[23,18,13,20]

Table 1: Summary of privacy threats and related works.

to analyze how customers' personal information is treated. Amazon declares they automatically collect information such as purchase history, IP address, e-mail address and miscellaneous browser information, in order to improve users' experience.[6] eBay also claims they collect similar information.[7] Indeed, the practice of gathering users' data is common (e.g., the *Magento* includes several extensions for dealing with customer relationship management[8]).

In [22], threat 1.1 in Table 1 is exposed for e-shops using PayPal. In this study, it was observed that 52% of the analyzed e-shops where sending product names, number of items and descriptions to PayPal. In addition, [22] also showed that PayPal leaked tracking information to Adobe's Omniture, including the referrer URL, which directly allows to link `C` and `M` (realizing threat 1.2 in Table 1). Moreover, note that in conventional e-shopping, risk 1.2 is always present, since `FN` and `DC` usually learn both `C` and `M` identities [5].

Threats in the checkout stage. In this stage, `C` specifies the payment information and shipping address. After applying fraud prevention techniques (e.g., reject purchases of more than a predefined price), `M` checks the promotions presented by `C`, if any, and forwards the payment information to `FN`. After validating the payment information (along with additional fraud prevention mechanisms), `FN` executes the payment. When checkout is completed, `M` updates `C`'s profile.

This stage handles most pieces of information; risks 1 to 6 and risk 13 of [5] (dealing with misuse, by `M` or an attacker, of the payment information) directly affect this stage. Namely, either `M` or `FN` may misuse `C`'s personal or payment information. Even if it is not misused by a dishonest entity, a *honest but curious* party may still pose a serious threat.

Concerning threat 2.1 in Table 1, as pointed out in [16], the current widely deployed 3-D Secure protocol, e.g., "Verified by Visa", "MasterCard SecuriCode", or "American Express SafeKey", requires a description of the transaction to be sent to `FN` (more exactly, the card issuer) in order for the cardholder to see and check it later. In particular, we know that some merchants leak product information to `FN` [22]. As to threat 2.2, in the protocol "Verified by Visa", `M` receives `C`'s PAN (Primary Account Number, i.e., the credit card number) [28].

A relevant example of threat 2.3 in Table 1, which may also imply threats 2.1 and 2.2, appears in [8]. From a large set of simply anonymized financial data (without names, addresses or obvious identifiers), [8] shows that it is possible to de-anonymize 90% of the individuals, if the data contain three items: price, when, and where. Note that it is very likely that the financial data collected by `FN` contain the three pieces of information. Price and time are directly known to `FN`. As for the location, for online purchases it may be deduced from IP addresses, shipping addresses, `M`'s information, etc. Worse yet, mobile-based payment would directly provide spatial coordinates through cellsite location [1,25].

[6] See "Amazon.com Privacy Policy" at `https://www.amazon.com/gp/help/customer/display.html?nodeId=468496`. Last access on January 13th, 2015.

[7] See eBay's "User Privacy Notice" at `http://pages.ebay.com/help/policies/privacy-policy.html`. Last access on January 13th, 2015.

[8] `http://www.magentocommerce.com`. Last access on June 27th 2015.

The concepts of *receiver privacy* and *value privacy* [15] formally capture the importance of the empirical analysis in [8]. As stated in [15], receiver privacy is maintained if "*the adversary cannot determine the receiver of a transaction, as long as this is issued by a non-compromised sender*", and value privacy is maintained if "*the adversary cannot determine the value of a transaction between two non-compromised users*". Ignoring value privacy implies leaking the price of a transaction, which significantly eases de-anonymization [8]. Ignoring receiver privacy leads to linking `C` and `M`, which may represent a privacy violation by itself (i.e., when buying sensitive products or services), or may lead to a privacy violation enabling re-identification of `C` from aggregated metadata.

Moreover, the payment information processed by financial entities includes card and account numbers, identifiers that persist across online and offline platforms and systems (unlike, e.g., cookies). This further implies that financial entities possess very sensitive information that paves the way to link purchases with payment transactions and perform behavioral analysis over customers' data [22]. Finally, fraud prevention is very relevant in the payment phase. It is the main mechanism that merchants and financial entities employ to prevent losses, which are far from negligible [1,2]. However, as pointed out in [1] new trends in these fraud prevention may pose a serious threat to privacy, like incorporating geolocation from mobile phones or information from social networks.

Threats in the delivery stage. Once `M` receives the payment, it delivers the purchased goods to `C`. For digital goods, the files are sent via Internet, and using anonymizing networks [9] is a robust way to protect privacy. For physical goods, these will be shipped through some delivery company `DC` to the shipping address specified by `C` to `M` during checkout (thus, realizing threat 3.1 in Table 1). Also, as pointed out in [5], depending on the information available to `DC`, it may pose additional privacy threats. In the real world, the delivery company `DC` at least learns both `C`'s and `M`'s addresses (threat 3.2 in Table 1), which allows it to link them, and may also learn other data, such as product related information.

However, preventing `M` (or other entities) from learning `C`'s physical address and `DC` to learn both `C`'s and `M`'s addresses is costly. Probably, physical mix networks are the most privacy respectful option [4]. Alternatively, Post Office boxes or equivalent delivery methods offer an intermediate solution between complexity and privacy, as it reveals a nearby location instead of `C`'s address.

Threats in the completion stage. After receiving the purchased items, `C` verifies that everything is correct, checking the debited amount, the received items, etc. If `C` is satisfied, the purchase is completed. If some error is detected, `C` may initiate a complaint. The situation is more complicated for purchases through e-shopping platforms (e.g., Amazon) rather than directly with the merchant; in this case, although it is usually recommended to first contact the merchant[9], it may be necessary for the e-shopping platform to mediate. In these situations, the privacy risks described for the previous stages will also be present, since `C` may need to provide product or payment information, or her contact information.

[9] See `https://payments.amazon.com/help/5968`. Last access on June 29th, 2015.

Additionally, whichever the final result is, `C` may provide online feedback about `M`, for other customers to decide whether or not to buy from him; in some platforms, such as eBay, `M` may also evaluate `C`. Concerning the possibility of leaving feedback, [13] shows how insufficient privacy controls may lead to serious privacy threats. Indeed, it is possible to infer the purchase history of a specific user by correlating the feedback she has received with the feedback received by the sellers with whom she has interacted. Also, it is possible to perform a *category attack* to obtain a list of the people that has bought an item of a specific type (e.g. guns). Other attacks explained in [13] include a *broad profiling attack* and a *side-information attack*, which also pose a serious threat to buyers (even enabling third parties to compromise their privacy in the case of the side-information attack). In a related context, [18] explains how to identify users in the Netflix database from little external information. All these attacks are realizations of threat 4.1 in Table 1. A more general model of the privacy threats related to recommendation systems is described in [23,20].

4 Proposals for Privacy-Preserving E-Shopping

There exists a reasonable amount of work addressing privacy flaws in different e-shopping subprocesses. Probably, most of the effort has been dedicated to payment systems. See, e.g., [12,14,3] for credit-card based systems protecting customers' privacy in different manners, or Bitcoin [17] and other cryptocurrencies for e-cash based systems. Private purchase systems are presented in [24], and iPrivacy [26] uses a proxy for sanitizing sensitive data from purchase orders. Fraud prevention has also been analyzed in the case of micropayments [6] and for the Bitcoin system as a means to prevent double-spending [11]. Also, privacy preserving marketing systems have received some attention [7,21]. Finally, [4] describes a physical mix network for physical objects delivery, and [26] proposes to use intermediate depots for protecting the customers' address.

However, as this summary of related privacy preserving systems shows, current proposals focus on specific stages of e-shopping. Nevertheless, if only one stage is protects privacy, attackers will just move to another one for breaking it. Moreover, solutions providing a subset of the functionality in current industry systems will probably be rejected by the main entities in the ecosystem, since all the mentioned features provide important usability and economic benefits.

5 Discussion

The process of an e-shopping transaction is complex, involving many entities, stages (Section 2) and features. In Section 3, we have outlined the privacy issues present at each stage. As a whole, the risks derived from those threats increase with the integration of all the mentioned processes in a complete infrastructure, as the *attack surface* increases as result of the greater complexity, amount and

variety of the information required to complete an e-shopping operation. However, only solutions for specific parts of the system have been proposed, putting aside the remaining components and leaving the overall design vulnerable.

The realization of any of the previous risks would imply serious threats to the privacy of online shopping customers. Not protecting against these threats may allow third parties to obtain customers' sensitive information. But privacy is not just a theoretical concern. As it has been empirically observed, when correctly informed about the subject, customers prefer privacy preserving online shops [27]. Moreover, customers are even willing to pay additional fees or higher prices for privacy preserving systems. For companies, this preference for privacy preserving alternatives is yet another incentive to address these issues.

To prevent rejection from the industry due to important features being ignored, we argue that first, global infrastructures promoting privacy and spanning the overall process should be devised. Second, proposals for each subsystem should be integrated within it, preventing information leaks. Finally, once every subsystem is implemented within the global architecture, a practical, comprehensive and privacy preserving e-shopping solution would have been achieved.

Thus, a central question is whether such a privacy supporting and feature-comprehensive solution is possible, being a central challenge in this scenario to find a good balance to satisfy both customers and service providers. An approach to achieve this is to employ the cryptographic primitives put forward in state-of-the-art Privacy Enhancing Technologies. A successful combination of these techniques would enable a robust solution, compatible with the complexity of the trust sharing features required by the e-shopping infrastructure.

Acknowledgements. This work was supported by project S2013/ICE-3095-CM (CIBERDINE) of the Comunidad de Madrid and MINECO TIN2010-19607, TIN2012-30883, TIN2014-54580-R. The work of Seung Geol Choi was supported in part by the Office of Naval Research under Grant Number N0001415WX01232. The work of Moti Yung was done in part while visiting the Simons Institute for Theory of Computing, UC Berkeley. The work of Jesus Diaz was done in part while visiting the Network Security Lab at Columbia University.

References

1. Ross J. Anderson. Risk and privacy implications of consumer payment innovation. `http://www.cl.cam.ac.uk/~rja14/Papers/anderson-frb-kansas-mar27.pdf`, 2012.
2. Ross J. Anderson, Chris Barton, Rainer Böhme, Richard Clayton, Michel van Eeten, Michael Levi, Tyler Moore, and Stefan Savage. Measuring the cost of cybercrime. In *WEIS 2012, Germany, 25-26 June, 2012*, 2012.
3. Elli Androulaki and Steven M. Bellovin. An anonymous credit card system. In *TrustBus*, pages 42–51, 2009.
4. Elli Androulaki and Steven M. Bellovin. APOD: Anonymous Physical Object Delivery. In *Privacy Enhancing Technologies*, pages 202–215, 2009.
5. Giannakis Antoniou and Lynn Margaret Batten. E-commerce: protecting purchaser privacy to enforce trust. *Electronic Commerce Research*, 11(4):421–456, 2011.

6. Matt Blaze, John Ioannidis, and Angelos D. Keromytis. Offline micropayments without trusted hardware. In *Financial Cryptography*, 2001.
7. Liqun Chen, Alberto N. Escalante, Hans Löhr, Mark Manulis, and Ahmad-Reza Sadeghi. A privacy-protecting multi-coupon scheme with stronger protection against splitting. In *Financial Cryptography*, pages 29–44, 2007.
8. Yves-Alexandre de Montjoye, Laura Radaelli, Vivek K. Singh, and Alex . Pentland. Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, 347(6221):536–539, January 2015.
9. Roger Dingledine, Nick Mathewson, and Paul F. Syverson. Tor: The second-generation onion router. In *USENIX Security Symposium*, 2004.
10. The Eurostat. E-commerce by individuals and enterprises.
11. Ghassan O. Karame, Elli Androulaki, Marc Roeschlin, Arthur Gervais, and Srdjan Capkun. Misbehavior in bitcoin: A study of double-spending and accountability. *ACM Trans. Inf. Syst. Secur.*, 18(1):2, 2015.
12. Steven H. Low, Nicholas F. Maxemchuk, and Sanjoy Paul. Anonymous credit cards and their collusion analysis. *IEEE/ACM Trans. Netw.*, 4(6):809–816, 1996.
13. Tehila Minkus and Keith W. Ross. I know what you're buying: Privacy breaches on ebay. In *PETS 2014, Amsterdam, July, 2014.*, 2014.
14. Ian Molloy, Jiangtao Li, and Ninghui Li. Dynamic virtual credit card numbers. In *Financial Cryptography*, pages 208–223, 2007.
15. Pedro Moreno-Sanchez, Aniket Kate, Matteo Maffei, and Kim Pecina. Privacy preserving payments in credit networks: Enabling trust with privacy in online marketplaces. In *NDSS 2015, San Diego, USA*, 2015.
16. Steven J. Murdoch and Ross J. Anderson. Verified by Visa and MasterCard SecureCode: Or, how not to design authentication. In *Financial Cryptography*, 2010.
17. Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2009.
18. Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (S&P 2008), 18-21 May 2008, Oakland, California, USA*, 2008.
19. U.S. Dept. of Commerce. The 2nd quarter 2013 retail e-commerce sales report.
20. Javier Parra-Arnau, David Rebollo-Monedero, and Jordi Forné. Optimal forgery and suppression of ratings for privacy enhancement in recommendation systems. *Entropy*, 16(3):1586–1631, 2014.
21. Kurt Partridge, Manas A. Pathak, Ersin Uzun, and Cong Wang. Picoda: Privacy-preserving smart coupon delivery architecture, 2012.
22. Sören Preibusch, Thomas Peetz, Günes Acar, and Bettina Berendt. Purchase details leaked to PayPal (Short Paper). In *Financial Cryptography*, 2015.
23. Naren Ramakrishnan, Benjamin J. Keller, Batul J. Mirza, Ananth Grama, and George Karypis. Privacy risks in recommender systems. *IEEE Internet Computing*, 5(6):54–62, 2001.
24. Alfredo Rial. *Privacy-Preserving E-Commerce Protocols.* PhD thesis, Arenberg Doctoral School, KU Leuven, 2013.
25. Norman M. Sadeh. *M-Commerce: Technologies,Services,and Business Models.* John Wiley & Sons, Inc., New York, NY, USA, 2002.
26. S. Stolfo, Y. Yemini, and L. Shaykin. Electronic purchase of goods over a communications network including physical delivery while securing private and personal information of the purchasing party, November 2 2006. US Patent App. 11/476,304.
27. Janice Y. Tsai, Serge Egelman, Lorrie Faith Cranor, and Alessandro Acquisti. The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, 22(2):254–268, 2011.
28. Visa. Verified by Visa – acquirer and merchant implementation guide, 2011.