Collaborative Filtering on the Blockchain: A Secure Recommender System for e-Commerce

Emergent Research Forum papers

Remo Manuel Frey
ETH Zurich
rfrey@ethz.ch

Dominic Wörner
MIT Media Lab
domwoe@mit.edu

Alexander Ilic
University of St. Gallen
alexander.ilic@unisg.ch

Abstract

In collaborative filtering approaches, recommendations are inferred from user data. A large volume and a high data quality is essential for an accurate and precise recommender system. As consequence, companies are collecting large amounts of personal user data. Such data is often highly sensitive and ignoring users' privacy concerns is no option. Companies address these concerns with several risk reduction strategies, but none of them is able to guarantee cryptographic secureness. To close that gap, the present paper proposes a novel recommender system using the advantages of blockchain-supported secure multiparty computation. A potential customer is able to allow a company to apply a recommendation algorithm without disclosing her personal data. Expected benefits are a reduction of fraud and misuse and a higher willingness to share personal data. An outlined experiment will compare users' privacy-related behavior in the proposed recommender system with existent solutions.

Keywords

Recommender system, collaborative filtering, e-commerce, blockchain, privacy, security.

Introduction

Today, recommender systems are everywhere in people's daily life and support them to make decisions in time. Especially in e-commerce, a reliable and efficient recommender system is essential. Vendors quickly realized the power and the importance for their own business. Over the last two decades, the usage of such systems changed from a trial balloon to a serious business tool (Sivapalan et al. 2014). There is an increasing expectation of user-tailored recommendations, often referred to as mass personalization (Kumar 2007). Several marketing studies proved that personalization improves the revenue thereby the satisfaction of the customers increases (Chen and Hsieh 2012; Smutkupt et al. 2010; Zhang and Wedel 2009). As a consequence, companies are creating detailed user profiles for a better understanding of customer behavior, needs, and habits. Unfortunately, privacy concerns prevent users to share data generously with interested companies, even if the quality of the recommendations would be improved. They mainly fear fraud and misuse of their data and a loss of control (Van Dyke et al. 2007). A novel blockchain-based approach (Zyskind et al. 2015a) is able to cryptographically guarantee the proper usage of personal data. The core component is a decentralized peer-to-peer network that allows storing encrypted data in a tamper-proof way and runs secure computations while no one but the data owner has access to the raw data. In the present paper, we propose to implement that approach in recommender systems. In doing so, a potential customer is able to allow a company to apply a recommendation algorithm without disclosing her profile. She never loses control of her data and is able to terminate the business relationship at any time. Fraud and misuse is no longer possible, because the involved company never gets the raw data. We speculate that such a system could become a standard for future recommender systems.

Literature Review

Recommender Systems and Data Privacy

A good definition of recommender systems in literature is provided by Encyclopedia of Machine Learning (Melville and Sindhwani 2010): “The goal of a recommender system is to generate meaningful recommendations to a collection of users for items or products that might interest them”. There are two well-known approaches to accomplish that goal: collaborative filtering (Goldberg et al. 1992) and content-based filtering (Pazzani 1999). Content-based Filtering tries to match knowledge of a single user with knowledge of available items while collaborative filtering tries to build recommendations based on knowledge of a collection of users. In order to get the advantages of both approaches, recommender systems in practice are typically a hybrid (Burke 2002). However, a content-based system may run on customer’s computer or mobile device. Sharing personal data with a company is not required and thus, a secure recommender system is easily achievable. In contrast, the case of collaborative filtering is different because a big collection of customers has to be considered to create one single recommendation. As a consequence, companies maintain huge databases for user data. They collect data like shopping history, shopping frequency, basket size, and redemption rate of coupons. ‘Big Data’ in e-commerce is reality today. The downside contains several dimensions of concerns: unauthorized secondary use (internal and external), improper access, and collection in general (Smith et al. 1996). Until now, there is no adequate solution in practice.

Data Privacy

Companies gather user data to create individual profiles and use it to predict consumer needs, favorite products, and preferred services. Moreover, sensors from the ‘Internet of Things’ starts to infuse people’s everyday life and companies are highly interested to receive such data from their customers. Sharing private data with a company might be a benefit for both – the company and the customer. Obviously, the nature of gathering personal data provoke questions about privacy protection. Who collects data and why? Is it protected against unauthorized access? Researchers investigated several factors, which have an influence on consumers’ privacy concerns, such as perception of risk (Mort et al. 2006), desire for control (Van Dyke et al. 2007), lack of trust (Roussos and Moussouri 2004), Anxiety and loss of comfort (Earp et al. 2005). Companies address these concerns with a wide arsenal of actions. For instance, they present a privacy policy to reduce the perception of risk and improve the transparency; and they provide an opt-out option as a kind of customer empowerment. However, none of these risk reduction instruments is able to cryptographically guarantee the anonymity of a customer, the secureness, or controllability of her data profile.

Privacy-preserving Computation

There has been a lot of effort to ease the tension between data mining and privacy. In principle there are two general approaches to store and mine data while preserving the privacy of individual users. On the one hand there is fully homomorphic encryption (Gentry 2009), which allows running computations on encrypted data and returning encrypted results on a central server. However, even for simple computations on small data the computational overhead is significant. Complex computations on large data sets, as typically needed for recommender systems, remain prohibitive. On the other hand there is secure multiparty computation (Lindell and Pinkas 2009). In this scheme data are encrypted, split into pieces, and shared among a distributed network of nodes. Computations are performed interactively and collaboratively on the network. Thereby, individual nodes do not get access to meaningful raw data, but only on encrypted shards. It is however essential that individual nodes are truly controlled by different stakeholders that will not collude. While a greater number of nodes would increase the protection against collusion, the performance decreases dramatically because of the increased communication complexity.

Blockchain-supported Secure Multiparty Computation

A blockchain is an open, immutable, append-only transaction log replicated among a network of nodes. Each block of transactions cryptographically references its predecessor and new blocks have to provide proof-of-work in order to be acknowledged by other participants in the network. Such a blockchain builds the basis of Bitcoin, the first peer-to-peer electronic cash system. Indeed every open (permissionless) blockchain is intertwined with a token of value i.e. a (crypto-) currency in order to provide economic incentives to follow the underlying protocol and extend the blockchain. The main features of a blockchain, an immutable public log, and a programmable token of value, have been used to advance secure multiparty computation systems in terms of fairness and operational efficiency (Andrychowicz et al. 2014; Bentov and Kumaresan 2014; Kiayias et al. 2015; Kumaresan and Bentov 2014). Enigma (Zyskind et al. 2015a, 2015b) implements those advancements in order to provide an open decentralized network for encrypted data storage and secure multiparty computations. Identity, access and contract management is facilitated by the underlying protocol. Private contracts provide the programming interface to access private and public data and to specify the computations. Thereby end users can permit and audit the usage of their data in fine granularity. Moreover they can revoke the permission at any time.

Proposed Solution

To sum up the literature review, recommender systems bring benefits to customers and companies as well. But, these systems rely on highly sensitive user data and ignoring users' privacy concerns is no option. Existent approaches are not able to cryptographically guarantee the right on privacy. To address that gap, we outline a secure recommender system using the advantages of blockchain-supported secure multiparty computation. Technical Insights of the underlying system (hereinafter referred to as 'Blockchain System' or 'BS' respectively) are described in the paper of Zyskind et al. (2015b). Customers are able to receive personalized recommendations without disclosing their data to anyone. What sounds like a contradiction, is becoming reality. We explain the interaction between customer, company, and system. To the best of our knowledge, a recommender system like this was never before described in literature. Our proposed solution is presented in three subsections.

Data Storage

A customer can use the BS to store personal data relevant for recommender systems. For instance, demographic attributes, favorite products, habits, special needs, and digital inventories. The data is fully encrypted and not accessible without explicit permission of the customer. More interesting data for recommender systems is typically collected by the companies itself, because their data reflects the real shopping behavior, not just an intention or consumer characteristics. For instance shopping history, and credit card data. To prevent hacker attacks and other risks, such data is stored in the BS (1) and not in the company anymore, as shown in Figure 1. The BS allows to handout such data on a trusted and encrypted way to the customer (2, 3). That means full transparency about companies' data collection activities for the customer. The data is visible for her at any time.

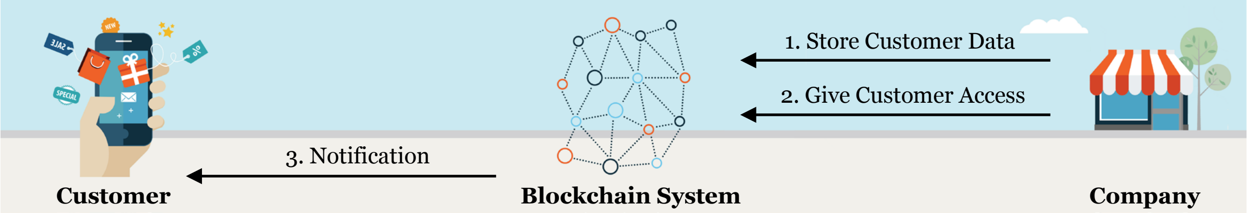

Figure 1. Data storage and handout.

Data Agreement Process

A company may use the BS for transparently defining contracts about how and what customer data are used in their recommender algorithms, as shown in Figure 2. First, the contract is stored in the BS (1). A new customer can get a company's contract by requesting it from the BS (2, 3). If the customer agrees with the provided contract, she can give access (4). Access means that the company is allowed to read and/or compute what is defined on the contract. The customer keep full control on her data and is able to terminate the contract at any time. A notification is sent to the company (5). A company may offer two kinds of incentives for sharing data with a company. First, sharing data can be set as a precondition to get personalized recommendations. Second, companies can pay money or offer discounts to get the desired data. As side effect, the well-known cold-start problem could be overcome.

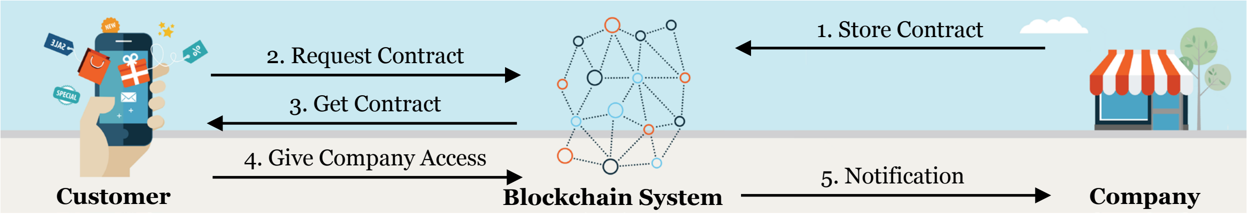

Figure 2. Data agreement process.

Collaborative Filtering

The strength of the BS is the possibility to do computations without disclosing the input data. As shown in Figure 3, the customer wants to get a recommendation for her (maybe confidential) shopping list (1). The company receives a notification (2). Since the company has now contracts with thousands of other people, their data can be used as input variables for a collaborative filtering algorithm to find out the optimal recommendation (3). Note that the company neither see customer's shopping list, nor the data from the thousand other people, nor the result because the whole computation is done in the BS. Finally, the customer receives a notification and gets secure access to the resulting recommendation (4).

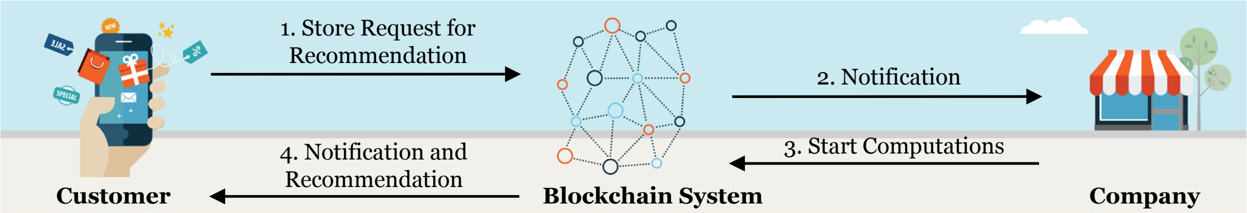
1. Store Request for Recommendation
2. Notification
3. Start Computations
4. Notification and Recommendation
Customer
Blockchain System
Company

Figure 3. Collaborative filtering.

Research Methodology

We develop a prototype of the presented recommender system. Besides the technical challenges, the research focus lies on the impact of the system on customers' privacy-related behavior. We will test the following hypothesis: *Compared to existent risk reduction strategies, the usage of blockchain technology reduces customer's perceived risks when sharing personal data in recommender systems.*

There are several existing risk reduction strategies. We intend to compare our approach with two well-established ones: customer empowerment (Van Dyke et al. 2007) and information transparency (Awad and Krishnan 2006). To test our hypothesis, we conduct an experiment with a questionnaire and form one test group per strategy. 'Strategy' is the independent variable. First, we explain to each group under which strategy their personal data from the questionnaire is handled. Then, we display the questionnaire where we ask all participants to enter several personal data items, namely from different levels of confidentiality (Victor et al. 2016): non-sensitive attributes (e.g. favorite color), sensitive attributes (e.g. income, illnesses), quasi identifiers (e.g. gender, age), and explicit identifiers (e.g. name, phone number). All answers are marked as optional. The number of entered data items per confidential level is used as dependent variable. We expect that participants in the blockchain group provide significantly more data items than in the other groups because the perceived risk is expected to be significantly lower. The results of the experiment will be analyzed using a one way independent ANOVA.

Discussion and Conclusion

Recommender systems may have a strong influence on the profitability of a company and they are often an integral part of companies' sales strategy. How to deal with customer profiles with respect to people's right on privacy is still a challenge in e-commerce. To the best of our knowledge, for the first time a solution is described which cryptographically guarantees customer's privacy in recommender systems. We outline a secure and powerful system, which has the potential to overcome customers' fear of fraud and misuse of their data. A secure handling of sensitive data is crucial for motivating people to share their consumption data and enables companies to create personalized recommendations. Moreover, comprehensive data access leads to new opportunities for researchers as well.

There are some limitations on the proposed system. First, we admit that the system does not prevent companies to collect customer data without permission. Second, manipulation of a fully anonymized recommender system like ours could be getting easier. A typical fraud is conducted by creating dummy profiles with the aim to manipulate the desirability of items and products. Either the rating of the own products is raised (push attack) or the rating of competing products is lowered (nuke attack). Third, Massa and Avesani (2007) proposed a network of trust to overcome the problem of data sparsity. Again, if the customers are completely anonymous, such concepts are difficult to realize. Finally, the solution has still to prove its computational performance. While its virtual machine is in principle able to do execute all kinds of computations, complex algorithms of modern recommender systems and large amounts of data might still be prohibitive. However, increasing power of computers and tailored implementations of algorithms will mitigate these problems. Until then, our system can at least provide limited recommendations based on simple algorithms for the privacy-concerned user who wouldn't give up her data for recommendations.

REFERENCES

Awad, N. F., and Krishnan, M. S. 2006. "The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiled Online for Personalization," *MIS Quarterly* (30:1), pp. 13–28.

Burke, R. 2002. "Hybrid Recommender Systems: Survey and Experiments," *User Modeling and User-Adapted Interaction* (12), pp. 331–370.

Chen, P. T., and Hsieh, H. P. 2012. "Personalized Mobile Advertising: Its Key Attributes, Trends, and Social Impact," *Technological Forecasting and Social Change* (79:3), Elsevier Inc., pp. 543–557.

Van Dyke, T. T., Midha, V., and Nemati, H. 2007. "The effect of consumer privacy empowerment on trust and privacy concerns in e-commerce," *Electronic Markets* (17:1), pp. 68–81.

Earp, J. B., Anton, A. I., Aiman-Smith, L., and Stufflebeam, W. H. 2005. "Examining Internet privacy policies within the context of user privacy values," *IEEE Transactions on Engineering Management* (52:2), pp. 227–237.

Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. 1992. "Using collaborative filtering to weave an information tapestry," *Communications of the ACM* (35:12), pp. 61–70.

Kumar, A. 2007. "From mass customization to mass personalization: A strategic transformation," *International Journal of Flexible Manufacturing Systems* (19:4), pp. 533–547.

Massa, P., and Avesani, P. 2007. "Trust-aware recommender systems," *Proceedings of the 2007 ACM conference on Recommender systems RecSys 07* (20), pp. 17–24.

Melville, P., and Sindhwani, V. 2010. "Recommender Systems," *Encyclopedia of Machine Learning*, Springer Science & Business Media.

Mort, G. S., Drennan, J., Sullivan, G., and Previte, J. 2006. "Privacy, Risk Perception, and Expert Online Behavior: An Exploratory Study of Household End Users," *Journal of Organizational and End User Computing* (18:1), pp. 1–22.

Pazzani, M. J. 1999. "A framework for collaborative, content-based and demographic filtering," *Artificial Intelligence Review* (13:5), pp. 393–408.

Roussos, G., and Moussouri, T. 2004. "Consumer perceptions of privacy, security and trust in ubiquitous commerce," *Personal and Ubiquitous Computing* (8:6), Springer-Verlag London Ltd, pp. 416–429.

Sivapalan, S., Sadeghian, A., Rahnama, H., and Madni, A. M. 2014. "Recommender systems in e-commerce," in *World Automation Congress Proceedings*, pp. 179–184.

Smith, H. J., Milberg, S. J., and Burke, S. J. 1996. "Information Privacy: Measuring Individuals' Concerns about Organizational Practices," *Management Information Systems Quarterly* (20:2), pp. 167–196.

Smutkupt, P., Krairit, D., and Esichaikul, V. 2010. "Mobile Marketing: Implications for Marketing Strategies," *International Journal of Mobile Marketing* (5:2), pp. 126–139.

Victor, N., Lopez, D., and Abawajy, J. H. 2016. "Privacy models for big data: a survey," *International Journal of Big Data Intelligence* (3:1), pp. 61-75.

Zhang, J., and Wedel, M. 2009. "The Effectiveness of Customized Promotions in Online and Offline Stores," *Journal of Marketing Research* (46:2), pp. 190–206.

Zyskind, G., Nathan, O., and Pentland, A. 2015a. "Decentralizing privacy: Using blockchain to protect personal data," *Proceedings - 2015 IEEE Security and Privacy Workshops*, pp. 180–184.

Zyskind, G., Nathan, O., and Pentland, A. 2015b. "Enigma: Decentralized Computation Platform with Guaranteed Privacy," *arXiv:1506.03471*.