Reusable Defense Components for Online Reputation Systems

Johannes Sänger, Christian Richthammer, Artur Rösch, Günther Pernul

Reusable Defense Components for Online Reputation Systems

Johannes Sänger, Christian Richthammer, Artur Rösch, and Günther Pernul

Department of Information Systems
University of Regensburg
Regensburg, Germany
{firstname.lastname}@wiwi.uni-regensburg.de
http://www-ifs.uni-regensburg.de

Abstract. Attacks on trust and reputation systems (TRS) as well as defense strategies against certain attacks are the subject of many research papers. Although proposing valuable ideas, they all exhibit at least one of the following major shortcomings. Firstly, many researchers design defense mechanisms from scratch and without reusing approved ideas. Secondly, most proposals are limited to naming and theoretically describing the defense mechanisms. Another issue is the inconsistent denomination of attacks with similar characteristics among different researchers. To address these shortcomings, we propose a novel taxonomy of attacks on TRS focusing on their general characteristics and symptomatology. We use this taxonomy to assign reusable, clearly described and practically implemented components to different classes of attacks. With this work, we aim to provide a basis for TRS designers to experiment with numerous defense mechanisms and to build more robust systems in the end.

Keywords: Trust, online reputation, reputation systems, attacks, taxonomy, components, reusability

1 Introduction

Electronic marketplaces like eBay and Amazon have greatly facilitated transaction processes between entities on the Internet. This provides many benefits but at the same time also poses significant challenges. One of the fundamental problems in electronic marketplaces is that, unlike in traditional face-to-face transactions, buyers do neither get a complete picture of a product's actual quality nor do they know about the trustworthiness of the particular seller. To address this, trust and reputation systems (TRS) have become important elements for the decision making process in this mostly anonymous environment. According to a recent study carried out by Diekmann et al. [2], sellers with better reputation are able to obtain higher prices and an increased number of sales. On the one hand, this can encourage good behavior because users seek good reputation to benefit from it. But on the other hand, TRS are likely to face an increasing amount of attacks by malicious users who try to gain unfair advantages by manipulating

the reputation system through specific behavior [6]. Therefore, it is fundamental for the providers to use TRS that are robust against all kinds of attacks that could lead to deceptive reputation scores and trust.

In order to be able to cover every possible attack scenario, we firstly develop a taxonomy of attacks in electronic marketplaces. On the highest level, we distinguish between attacks performed as a seller (*seller attacks*) and attacks carried out in the role of the buyer (*advisor attacks*). Then, we identify defense mechanisms for different types of attacks by assigning reusable TRS components that can be employed to extend the functionality of the computation engine. These components are provided in the form of both a conceptual description and fully implemented reusable web-services in the component repository[1] introduced by Sänger and Pernul [10]. The additional attack view on TRS components constitutes an important extension to the yet largely functional view. We argue that the assignment of TRS components to attack types not only supports the development of more reliable and robust TRS with already existing components but also helps to identify weaknesses that have not been addressed so far.

The remainder of the paper is organized as follows. Firstly, we give an overview of the general problem context of our work in Section 2. Thereby we delineate the research gap we discovered and define the objectives of our proposal. In Section 3, we introduce our novel taxonomy of attack types on TRS. We use this taxonomy in Section 4 to assign TRS components to the different classes of attacks. At the same time, we point out how the outcomes of this allocation are described in clearly structured attack profiles and integrated in the knowledge repository. In Section 5, we discuss our findings before we conclude in Section 6.

2 Problem Context and Related Work

As opposed to traditional face-to-face interactions, the "universe of strangers" [1] found in electronic marketplaces makes it hard to determine the trustworthiness of an actor. This is due to insufficient information as entities commonly never have transacted with each other before. The problems resulting from the lack of information can be mitigated through TRS, which have become a widely adapted element for the decision making process in online environments. To establish a common understanding, we firstly point out related work on attacks on TRS. After that, we briefly describe the reusable TRS repository whose components we map against our attack classes. This leads us to the research gap we address in this paper.

2.1 Attacks on Trust and Reputation Systems

TRS can be subject to attacks by their participating entities in various ways. Attacks may be dependent on the specific application scenario, influenced by the social environment underlying the reputation system, and performed by one

[1] http://trust.bayforsec.de/

single entity or by several colluding entities. Because of the increasing attention paid to attacks against TRS, several security analyses were carried out in recent years [3, 4, 6, 13]. The resulting proposals of attack taxonomies and formulations of challenges for robust TRS in turn motivated studies on defense strategies (for related surveys see [5] and [8]). As the various trust models are specifically designed to cope with certain attacks, they are not completely robust against various attacks in different settings. Therefore, security and robustness still remain the key challenges in the design and development of TRS.

2.2 Reusable Component Repository

Since most of the TRS described in literature use computation methods that are entirely built from scratch [12], well-established approaches are rarely considered. To foster reusability, Sänger and Pernul [10] proposed a hierarchical component taxonomy of computation engines along with a repository containing design knowledge both on a conceptual and an implementation level. On the conceptual level, they described each building block as a design pattern-like solution. On the implementation level, they provided fully implemented reusable components by means of web-services. The classes of the component repository were the result of the analysis of their generic process of reputation systems as well as various computation methods described in different surveys [7, 9, 11–13].

2.3 Research Gap

Apart from the component repository described before, further important steps toward reusability were made by Hoffman et al. [5] and Koutrouli and Tsalgatidou [8]. They conducted surveys on attacks and defense mechanisms and thus helped to collect the ideas for the research community. The main shortcoming of these surveys is that they are limited to naming and theoretically describing the defense mechanisms.

In this paper, we want to go one step further and employ the reusable computation components described by Sänger and Pernul [10] as defense mechanisms for attacks on TRS. The uniform format of their design pattern-like artifacts helps to establish clear guidelines for developing new defense mechanisms. Moreover, their fully implemented components by means of web-services allow researchers to experiment.

In a preparatory step, we aim to extend their repository by an attack view in which we systematically describe attack types with certain characteristics instead of basing the discussions on particular examples of attacks. While this helps to avoid the yet inconsistent denominations of some attacks (e.g. re-entry vs. whitewashing), it also makes our remarks more generic and extendable. Most importantly, we are then able to assign reusable computation components to entire classes of attacks instead of matching the same defense methods against numerous examples of attacks.

3 Taxonomy of Attacks on Trust and Reputation Systems

In this section, we introduce a novel attack taxonomy for electronic marketplaces in order to organize possible kinds of attacks. On the highest level, we distinguish between seller attacks and advisor attacks. In these major classes, we classify every attack type along two dimensions: attackers and behavior.

3.1 Seller Attacks vs. Advisor Attacks

In a common electronic marketplace, we have two parties: the buyer and the seller. In terms of TRS, both can take the role of the ratee (the one being rated, usually the seller) and the advisor (the one who provides a referral, usually the buyer).

To decide which seller to transact with, buyers rely on ratings of other buyers to evaluate the reputation of sellers. A seller that delivers an item as specified in the contract is referred to as an honest seller, whereas a seller that does not deliver an item as specified in the contract is called a dishonest or malicious seller. Note that the term "item" includes both physical and non-physical products as well as services. Seller attacks denote manipulations of the reputation system that one or more entities of an electronic marketplace perform in the role of the seller. The intention behind these manipulations is to be able to act as a malicious seller while maintaining a reputation profile that buyers would assess as honest. Even though cheating behavior from dishonest sellers (e.g. not delivering an item at all) can be sentenced by law, TRS should aim to prevent these actions from the first.

Advisor attacks, in contrast, are implemented by the rating parties. Since buyers can usually rate a seller's performance in a particular transaction, they are able to shape his reputation profile and thus act as advisors for other buyers. According to Jøsang and Golbeck [6], advisor attacks can be summarized under the term "unfair rating attacks" because they are based on one or several digital identities providing unfair ratings to other digital identities. These unfair ratings are used to manipulate the reputation profile of sellers – either boosting or vilifying it to an unjustified extent. As opposed to seller attacks, advisor attacks can generally not be sentenced by law.

3.2 Dimensions: Attackers and Behavior

Within the classes of seller and advisor attacks, our taxonomy systematizes attack types along the two dimensions: attackers and behavior.

Attackers The attackers dimension refers to the number and characteristics of the digital identities participating in an attack. Although seller attacks are typically performed by one single digital identity, some of them may also be performed by a colluding group of attackers. Depending on the trust model and identity management concept used by the reputation system, attackers may also create additional digital identities on their own in order to boost their leverage.

– One identity: An attacker performs all actions on his own, independently and without the help of other entities. Furthermore, he does not create any additional accounts but conducts the attack with one single digital identity.
– Multiple identities: In online environments, which are mostly anonymous, pseudonyms can generally be created with minimal costs. Hence, a malicious entity may easily acquire multiple digital identities with which he is able to create pseudo-referrals and boost his reputation in the system.
– Multiple entities: A group of attackers agrees to perform a joint attack. Typically, the damage caused by multiple colluding entities is considerably higher than by entities independently performing malicious actions.

Behavior The behavior dimension characterizes the actions of an attacker. Here, we differentiate attackers acting maliciously all the time from attackers alternating between malicious and honest actions.

– Consistent: Attackers act maliciously all the time and do not perform any honest actions.
– Inconsistent: Attackers perform both honest and dishonest actions. Thus, the dishonest actions can be used to gain higher profits, for instance, while the honest actions ensure that the reputation value is kept at a level that makes other users assess the attacker as honest.

4 Introducing an Attack View on the Component Repository

In this section, we show how we implemented the novel "attack view" on the component repository. Thereto, we firstly accomplish the assignment of attack classes and defense components. Secondly, we delineate how the taxonomy of attacks was integrated as part of the knowledge repository and linked to the computation components.

4.1 Assignment of Defense Components

Most research papers on defense mechanisms against attacks in TRS propose a variety of possible solutions in form of "unstructured" textual recommendations. In this work, in contrast, we assign reusable components. These components are not only implemented in a web-service but also clearly described in well-structured design pattern-like artifacts. In this way, a developer can directly make use of both the ideas and the web-services that can be integrated in existing reputation systems to extend their capabilities.

To accomplish the assignment, we analyzed the single classes of our taxonomy of attacks on TRS in electronic marketplaces introduced in the former section with regard to their general characteristics. Table 1 shows an excerpt of the results. The terms listed on the right side of the table reflect the unique component terms as used in the component repository. These components provide a range of different defense approaches that can be applied either alone or in combination.

Table 1. Excerpt of assignment table

Primary class	**Secondary class**	**Tertiary class**	**Component**
seller	one identity	consistent	Summation, Bayesian Probability, Average, Share (positive)
		inconsistent	Asymetric rating, Absolute time discounting, Relative time discounting, Age-based filter, Context similarity, Criteria similarity
...	...	...	...
advisor	multiple identities	consistent	Clustering filter, Subjective reliability
		inconsistent	Absolute time discounting, Beta-statistic filter, Clustering filter, Propagation discount, Relative time discounting, Subjective reliability

4.2 Implementation as Part of the Knowledge Repository

In the second step, we implemented the taxonomy of attacks as part of the knowledge repository[2].

Table 2. Example profile of one attack class, shortened

Attack Classes	Advisor attack: consistent (One identity)
Description	In a consistent advisor attack carried out by one identity, a single advisor consistently assigns deceptive ratings to transactions. This means consistently providing unfairly low ratings to honest sellers and/or consistently providing unfairly high ratings to dishonest sellers.
Examples	- Consistent ballot stuffing: The attacker provides unfairly high ratings toward other actors to increase their reputation. - Consistent bad mouthing: The attacker provides unfairly low ratings toward other actors to discourage their reputation. [...]
Solution	There are several ways to filter out unfair ratings made by single attackers. Detection/filtering mechanisms can broadly be divided into two groups: endogenous filtering/discounting and exogenous filtering/discounting. Endogenous discounting methods try to detect unfair ratings on the basis of their statistical properties. [...]
Pattern/web-service	– Beta-statistic filter – Clustering filter – Objective reliability (reputation) [...]
Literature	– Tavakolifard, M., Almeroth, K. A Taxonomy to Express Open Challenges in Trust and Reputation Systems. Journal of Communications, North America, 7, 7. 2012. [...]

[2] http://trust.bayforsec.de/ngot/index.php?section=knowledge_repository

To give a more detailed view on the single classes as well as the possible defense strategies, we described each block in a clearly structured "profile". Each profile contains a general description for that block, a number of example attacks, a solution (defense strategy) to that problem, hyperlinks to design patterns/web-services that can be used to implement the solution, and a list of relevant literature. All these profiles can be found online as part of the knowledge repository. Table 2 depicts an example profile for a *consistent* advisor attack based on *one identity*.

5 Discussion

Reviewing the assignment of our taxonomy of attacks and defense mechanisms, we made some interesting findings. In contrast to most surveys on attacks and defense mechanisms for TRS, we did not introduce a range of different attacks in this work but rather focused on the general characteristics and symptomatology of attacks such as the continuity and the number of attackers. We thereby found that many attacks that have been described as distinct challenges in literature are actually different manifestations of the same symptomatology. Consequently, defense mechanisms against specific characteristics of attacks may help to cover a variety of challenges.

Overall, the assignment of attack classes and computation components brings some valuable benefits:

- Developers not only gain solutions to challenges stemming from weaknesses against attacks in form of a short recommendation but find a clearly structured design pattern-like description of the exact problem, a solution to that problem, a generic code example and further literature. Moreover, they can directly make use of a web-service implementing that logic.
- Having a range of already implemented services, developers can experiment with different combinations of components to find the best solution for their specific problem, TRS and use case.
- Researchers are encouraged to use this clearly defined structure when developing new ideas and defense mechanisms, and make them available in form of both design patterns and web-services in the component repository.

6 Conclusion

Lots of research on attacks and defense strategies on TRS has been done in the past. In this paper we developed a novel taxonomy which, to the best of our knowledge, is the first taxonomy that can be used to describe all attacks that focus on the manipulation or exploitation of the reputation computation in e-commerce settings. We then identified defense mechanisms for different types of attacks by mapping reusable TRS components against classes of attacks. In this way, we not only support reputation system designers in the development of more reliable and robust TRS with already existing components but also help

to identity weaknesses that have not been addressed so far. Furthermore, our taxonomy is valuable for future research in that it provides a basis to describe attacks by their characteristics and symptomatology and contributes to a common understanding of attacks on TRS.

Acknowledgements The research leading to these results was supported by the "Bavarian State Ministry of Education, Science and the Arts" as part of the FORSEC research association.

References

1. Dellarocas, C.: Reputation mechanisms. In: Hendershott, T. (ed.) Handbook on Economics and Information Systems, pp. 629–660. Elsevier Publishing (2006)
2. Diekmann, A., Jann, B., Przepiorka, W., Wehrli, S.: Reputation formation and the evolution of cooperation in anonymous online markets. American Sociological Review 79(1), 65–85 (2014)
3. European Network and Information Security Agency: Reputation-based Systems: A Security Analysis (2007), European Network and Information Security Agency
4. Fraga, D., Bankovic, Z., Moya, J.M.: A Taxonomy of Trust and Reputation System Attacks. In: Proc. of the 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). pp. 41–50 (2012)
5. Hoffman, K., Zage, D., Nita-Rotaru, C.: A Survey of Attack and Defense Techniques for Reputation Systems. ACM Computing Surveys 42(1), 1–31 (2009)
6. Jøsang, A., Golbeck, J.: Challenges for Robust Trust and Reputation Systems. In: Proc. of the 5th International Workshop on Security and Trust Management (STM) (2009)
7. Jøsang, A., Ismail, R., Boyd, C.: A Survey of Trust and Reputation Systems for Online Service Provision. Decision Support Systems 43(2), 618–644 (2007)
8. Koutrouli, E., Tsalgatidou, A.: Taxonomy of Attacks and Defense Mechanisms in P2P Reputation Systems - Lessons for Reputation System Designers. Computer Science Review 6(2-3), 47–70 (2012)
9. Noorian, Z., Ulieru, M.: The State of the Art in Trust and Reputation Systems: A Framework for Comparison. Journal of Theoretical and Applied Electronic Commerce Research 5(2), 97–117 (2010)
10. Sänger, J., Pernul, G.: Reusability for Trust and Reputation Systems. In: Trust Management VIII: Proc. of the 8th IFIP WG 11.11 International Conference (IFIPTM). pp. 28–43 (2014)
11. Sherchan, W., Nepal, S., Paris, C.: A Survey of Trust in Social Networks. ACM Computing Surveys 45(4), 1–33 (2013)
12. Tavakolifard, M., Almeroth, K.C.: A Taxonomy to Express Open Challenges in Trust and Reputation Systems. Journal of Communications 7(7), 538–551 (2012)
13. Yao, Y., Ruohomaa, S., Xu, F.: Addressing Common Vulnerabilities of Reputation Systems for Electronic Commerce. Journal of Theoretical and Applied Electronic Commerce Research 7(1), 3–4 (2012)