Computers in Human Behavior 57 (2016) 274–291

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Benchmarking reputation systems: A quantitative verification approach

Amir Jalaly Bidgoly, Behrouz Tork Ladani*

Department of Computer Engineering, University of Isfahan, Isfahan, Iran

A R T I C L E I N F O

Article history:
Received 26 March 2015
Received in revised form
19 November 2015
Accepted 13 December 2015
Available online 28 December 2015

Keywords:
Reputation systems
Robustness
Quantitative verification
Benchmarking
RepSyFire

A B S T R A C T

Trust and reputation systems are classes of decision support tools which help detecting malicious behavior based on collecting ratings and opinions. Despite their advantages, these systems are vulnerable to some kinds of attacks in which the attacker can deceive the system using sequences of misleading behaviors. Robustness of reputation systems against these attacks are frequently investigated in the literature. However the existing works usually evaluate the robustness using a qualitative simulation method. Lack of a formal evaluation method and a quantitative measure of robustness make it hard to extend the results and to compare the systems precisely. This paper proposes a quantitative robustness measure for reputation systems based on a formal verification approach. Using the robustness measure and the verification method, a comprehensive benchmarking of a number of well-known reputation systems is presented which includes evaluation of the systems against basic and the worst case attacks. The results are used for ranking and classifying the systems. The studies show that robustness is not an absolute feature of a reputation model, but it also depends on the properties of the environment. The benchmarking results have been also used to indicate the proper environment for each class of systems/attacks.

1. Introduction

Trust and reputation systems (TRSs) are classes of decision support tools that are widely used in electronic communities. The objective of these systems is to compute the trustworthiness rank of each member of the community based on collecting ratings and opinions of the members. The trustworthiness rank is then used to filter dishonest members or select the best participant. TRSs are employed in many environments, including online service providers and e-Market places (Jøsang, Ismail, & Boyd, 2007; Sabater & Sierra, 2005; Wang & Vassileva, 2007), semantic web (Artz & Gil, 2007; Golbeck, 2006; Zhang, Chen, & Wu, 2006), social networking (Gomez Marmol, Gil Perez, & Martinez Perez, 2014), wireless communication (Yu, Shen, Miao, Leung, & Niyato, 2010; Fernandez-Gago, Roman, & Lopez, 2007), multi-agent systems (Huynh, Jennings, & Shadbolt, 2006; Ramchurn, Huynh, & Jennings, 2004), Ad-hoc networks (Cho, Swami, & Chen, 2011; Yih-Chun & Perrig, 2004; Zhang, 2011), and even intrusion detection systems (Fung, Zhang, Aib, & Boutaba, 2011; Pérez, Tapiador, Clark, Pérez, & Gómez, 2014). They also have gained lots of interest in online peer-to-peer interaction communities such as Amazon, eBay, Yahoo, YouTube, Yelp, and CouchSurfing.

Despite their wide application, TRSs may have vulnerabilities in which malicious attackers can exploit to perform sequences of misleading behavior with the aim of gaining unfair trust/reputation scores. A vulnerable TRS not only fail in filtering attackers, but also can be used by them to empower their attacks. TRSs should be designed in a robust way, i.e. they can function properly even in the existence of malicious attackers (Jøsang, 2012; Jøsang & Golbeck, 2009; Muller, Liu, Mauw, & Zhang, 2014). The proposals of new TRSs, typically do not contain comprehensive evaluation studies on the robustness of the proposed systems (Jøsang, 2012; Jøsang & Golbeck, 2009). The evaluation is normally limited to performing a number of simple case studies to show how the system meets the specified requirements.

Selecting the best fitted TRS for a given environment requires to conduct comprehensive evaluation and comparison between the existing systems. There are a number of researches in the literature with the aim of presenting comprehensive studies on different aspects of TRSs including the robustness of the systems against malicious attackers. However, the existing work suffer from some

* Corresponding author.
E-mail address: ladani@eng.ui.ac.ir (B. Tork Ladani).

http://dx.doi.org/10.1016/j.chb.2015.12.024
0747-5632/

common problems. The researches typically suffice to a qualitative robustness evaluation to check the existence of required defense mechanisms against known attacks. Also, those which aimed to propose quantitative approaches do not contain a precise and formal robustness measure. Beside, these work always choose simulation as the underlying evaluation method. Using the simulation approach causes the evaluation to be limited to known attacks, whereas verification as its alternative, can search all state space of the system to discover new and unknown attacks.

Although trust and reputation are two close concepts which are frequently used instead of each other, they should be distinguished. Trust is a subjective value that indicates the trustworthiness of an entity from the viewpoint of another entity, whereas reputation is a global value in the consequence of the whole community opinions towards an entity (Lopez, Roman, Agudo, & Fernandez-Gago, 2010; Wang & Vassileva, 2007). Trust systems (TSs) typically assign a trust value to each pair of entities by evaluating their direct (or indirect) experiences, while reputation systems (RSs) compute a single score for each entity based on aggregation of all entities experiences. The dissimilarities between TSs and RSs make them behave differently. For example, an RS is usually (not always) implemented within a centralized structure, while TSs are more used in distributed structures. From the robustness point of view, attackers may choose different malicious behavior against TSs and RSs. For instance, discrimination attack (Jøsang et al., 2007), that is the result of behaving differently between groups of entities, can be performed against TSs simpler than RSs. It can be argued that the robustness of RSs and TSs should be evaluated separately.

In this paper, we present a benchmarking approach to assess and compare the robustness of some selected, but famous reputation models. To do that, a novel robustness measure for reputation systems is proposed. The measure is based on a previously presented method for quantitative verification of RSs (Jalaly Bidgoly & Tork Ladani, 2015). We have also implemented the verification method as a tool named *RepSyFire* (Reputation Systems Verifier) which is used for benchmarking in this work. The performed benchmarking not only include verification and comparison of selected reputation models against basic attacks (such as malicious service provision and unfair rating), but also finds and compares the worst (i.e. the most powerful) existing attacks against the given RSs. The achieved results reveal the robustness value of the selected RSs. This way we are able to classify the RSs and to indicate the robustness rank of each one. The results show that robustness of a given RS varies in different environments in such a way that a robust RS in one environment may turn into a vulnerable system in another environment. The benchmarking results have been used to determine proper environments for the given RSs/attacks.

The contribution of this work can be highlighted as follows:

- Introducing a novel robustness measure.
- Verifying the robustness of a wide range of RSs using a formal verification approach (As far as we know it is the first try in this field).
- Comparing and exploring the strengths and weaknesses of the RSs in different environment and against different attacks. These results help the researchers and managers to select the right reputation model for each environment.
- Presenting the worst case analysis of RSs. The worst case analysis just can be performed using verification approaches, hence previous works that use simulation approach are not capable of performing this kind of analysis.
- Analyzing the impact of different parameters on robustness. For instance, here it is concluded that probabilistic selection is not a proper choice for environments that contain attackers.
- Ranking and classification of RSs.
- Introducing RepSyFire as an open source tool for verification of RSs which make the process of benchmarking clearer and easier.

Table 1
The list of atomic action and corresponding rewards.

	Cost	Reward	Total
U_{sp}	C_{REQ}	0	$-C_{REQ}$
P_m	C_{REQ}	R_{REQ}	$R_{Req}-C_{REQ}$
S_v	$C_{QTY}(v)$	R_{REQ}	$R_{REQ}-C_{QTY}(v)$
E	C_{ID}	0	$-C_{ID}$
N	0	0	0

Table 2
The Reward model in dishonest behavior evaluations.

Cost	Value	Reward	Value
C_{REQ}	10	R_{REQ}	10
$C_{QTY}(sat)$	9	α_{SLN}	0
$C_{QTY}(unsatisfactory)$	3		
C_{ID}	0		

The paper continues as follows: Next section reviews the related work on robustness evaluation of TRSs, in Section 3, first the underlying verification method for RSs is reviewed, and then the proposed robustness measure is introduced. Section 4 exhibits the modeling of the selected reputation systems. Sections 5 and 6 represent the results of benchmarking of RSs for the basic attacks and the worst case attacks respectively. Finally the paper ends in Section 7 with the conclusion.

2. Related work

The concepts of computational trust and reputation are closely related together in the literature. Hence, here in this section the related work on the validation of both trust systems and reputation systems are reviewed. There are a number of well-known attacks against TRSs in the literature (Hoffman, Zage, & Nita-Rotaru, 2009; Marmol & Perez, 2009; Singh & Kumar, 2011; Touceda, Sierra, Izquierdo, & Schulzrinne, 2012). These attacks include a wide range of malicious behaviors from simple behaviors such as selfishness and unfair rating to complicated ones like *RepTrap* and *RepHi* that have been addressed in (Feng, Zhang, Chen, & Fu, 2011; Yang, Feng, Sun, & Dai, 2008). There are lots of work on the general domain of evaluating TRSs. A big part of them are those which are limited to a single model or a specified environment. For instance, analyzing on-off attack in wireless sensor networks (Chen, Zhang, Liu, & Feng, 2010; Perrone & Nelson, 2006; Shi & Chen, 2012), evaluating eBay system (Dini & Spagnolo, 2009; Resnick, Zeckhauser, Swanson, & Lockwood, 2006; Resnick & Zeckhauser, 2002), Beta reputation system (Bidgoly & Ladani, 2013), and PageRank (Clausen, 2004) are just examples of these work.

There are limited researches that exclusively focus on robustness evaluation and comparison of TRSs. The aim of these works is just presenting the TRSs evaluation results and comparing their strengths and weaknesses with each other. The presented work by Hoffman et al. (Hoffman et al., 2009) is a pioneer and the most cited paper which classifies the existing TRSs, their attacks and their defense mechanisms. They have also analyzed the strengths and weaknesses of a selected number of TRSs in details. Likewise, some other classifications for TRSs and their attacks have been proposed by Jøsang (Jøsang & Golbeck, 2009) and Mármol (Gómez Mármol & Martínez Pérez, 2010; Marmol & Perez, 2009). They also suggested general solutions to protect against the attacks. Noorian et al. in (Noorian & Ulieru, 2010) also have defined a set of so called hard