Chapter 12

An Evaluation Framework for Reputation Management Systems*

Andrew G. West  
University of Pennsylvania, USA

Sampath Kannan  
University of Pennsylvania, USA

Insup Lee  
University of Pennsylvania, USA

Oleg Sokolsky  
University of Pennsylvania, USA

ABSTRACT

Reputation management (RM) is employed in distributed and peer-to-peer networks to help users compute a measure of trust in other users based on initial belief, observed behavior, and run-time feedback. These trust values influence how, or with whom, a user will interact. Existing literature on RM focuses primarily on algorithm development, not comparative analysis. To remedy this, the authors propose an evaluation framework based on the trace-simulator paradigm. Trace file generation emulates a variety of network configurations, and particular attention is given to modeling malicious user behavior. Simulation is trace-based and incremental trust calculation techniques are developed to allow experimentation with networks of substantial size. The described framework is available as open source so that researchers can evaluate the effectiveness of other reputation management techniques and/or extend functionality. This chapter reports on the authors’ framework’s design decisions. Their goal being to build a general-purpose simulator, the authors have the opportunity to characterize the breadth of existing RM systems. Further, they demonstrate their tool using two reputation algorithms (EigenTrust and a modified TNA-SL) under varied network conditions. The authors’ analysis permits them to make claims about the algorithms’ comparative merits. They conclude that such systems, assuming their distribution is secure, are highly effective at managing trust, even against adversarial collectives.

DOI: 10.4018/978-1-61520-682-7.ch012
INTRODUCTION

At the start of the network-age the client-server (centralized) model was the dominant topology. Trust in these servers was implicit and security measures focused on access control and user permissions. More recently, new network architectures and computing paradigms have emerged such as distributed systems, peer-to-peer (P2P) networks, and ad-hoc mobile computing. Frequently, all network nodes have the ability to both request services from and provide services to other users. This is inherently risky since decentralized models typically lack the notions of authenticity, reliability, and accountability that monolithic servers provide. Nonetheless, well-behaved decentralized systems are beneficial in comparison to their client-server counterparts. Advantages include increased service diversity, availability, scalability, and bandwidth.

Enforcing good behavior is the task of a trust management (TM) system. The seminal work of Blaze, Feigenbaum, and Lacy (1996) introduced the term -- their system consisted of using cryptographically delegated credentials and policies to specify static access control rights. In reputation management$^1$ (RM), rather than determining if a user has the authority/permission to do some action, we instead ask: Given permission, how do we expect a user to behave (i.e., what is his/her reputation)? A systems treatment of these expectations gives rise to a dynamic access control mechanism which is categorically different than that provided by TM. Reputation management is implemented by a RM system or reputation algorithm$^2$ (RA).

Almost universally, RAs work by using past behavior as a basis for future conduct. Transitive trust is often exploited, especially in the absence of prior interaction between two parties. To promote a well-intentioned network either bad behavior is punished, good behavior rewarded, or both. EigenTrust (Kamvar, Schlosser, & Garcia-molina, 2003) and Trust Network Analysis with Subjective Logic (TNA-SL) (Jøsang, 2001; Jøsang, Hayward, & Pope, 2006) are two RAs that will be given particular attention herein. For a broader survey of available systems, readers should review the work of Li and Singhai (2007). One should note that the need for RM is not confined to purely digital dealings. In fact, eBay manages one of the largest RM systems (Resnick, & Zeckhauser, 2001), pertaining to the exchange of physical commodities.

Research concerning RM has been focused on algorithm development with little attention given to quantitative comparative analysis between existing RAs (qualitative analyses are often seen, but we feel, insufficient). Tests on some systems, like EigenTrust, use briefly-described, proprietary, or closed-source simulators (Schlosser, Condie, & Kamvar, 2003). Others, like TNA-SL, opt for a more theoretical description with no evaluation results. In order to compare systems such as these and verify author’s claims, an objective simulator is needed. While network and P2P simulators exist, having the additional overhead of simulating Distributed Hash Tables (DHTs), latency, network hops, etc., in addition to trust calculation make their use computationally inappropriate. Furthermore, such simulators offer little abstraction, making the implementation of RAs inconvenient. Therefore, in this chapter we describe the construction of an evaluation framework specific to reputation management.

This chapter is organized as follows: We will begin by standardizing terminology and justifying our architecture of choice. Trace generation and simulation under this architecture will then be discussed. Next, evaluation metrics with regard to effectiveness and efficiency will be introduced. Then, test runs will be used to exemplify behavior and identify potential shortcomings in our design. Finally, future work will be noted and concluding remarks made.

---

$^1$ RM

$^2$ RA