Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 4 (2011) 1751–1760

Procedia Computer Science

International Conference on Computational Science, ICCS 2011

A Data-Driven Framework for Dynamic Trust Management

Olufunmilola Onolaja*, Georgios Theodoropoulos, Rami Bahsoon

School of Computer Science, The University of Birmingham, United Kingdom, B15 2TT

Abstract

Reputation and trust-based models have been used extensively in different application domains. These include large online communities such as eBay, Amazon, YouTube and ad-hoc and wireless sensor networks. Recently, the use of the models has gained popularity due to their effectiveness in providing trusted systems or networks. These models focus on online and historical data to determine the reputation of domain members. In this paper, we propose a novel approach for obtaining trust values by focusing not only on online and historical data but also possible future scenarios to anticipate events in the next time intervals. The data-driven framework is able to dynamically obtain and inject data to predict the future trust value of every identity in the system. The advantage of this proactive approach compared to other approaches is that informed decisions about the domain can be made before a compromise occurs.

Keywords: Reputation, Trust, Data-driven systems

1. Introduction

Reputation and Trust-based models (RTMs) have gained popularity over the years, borrowing ideas from game theory and Bayesian networks. RTMs are described as systems that provide mechanisms to produce a metric encapsulating reputation for each identity in a given application domains [1]. Generally, RTMs aim to provide information that allow nodes to distinguish between trustworthy and untrustworthy members. The models encourage members to cooperate through the use of incentives, and discourage maliciousness by punishment schemes such as isolation and service denial.

RTMs have been adopted in applications that rely on the cooperation of domain members in order for the application to function correctly. The models have been used extensively in various e-commerce and online communities such as YouTube, Amazon and eBay as described in Section 2. Some literatures also suggest their use in domains ranging from peer-to-peer (P2P) to mobile networks [2, 3, 4].

A common problem of RTMs is their vulnerability to *collusion attacks*, where two or more nodes can team up to behave maliciously. Incentive policies that are used in P2P networks to ensure cooperation between nodes are generally susceptible to collusion attacks as well. Traditionally, the models rely on recommendations that are based on past interactions between the members provided by the same members to decide on the reputation of one another.

*Corresponding author

Email addresses: O.O.Onolaja@cs.bham.ac.uk (Olufunmilola Onolaja), G.K.Theodoropoulos@cs.bham.ac.uk (Georgios Theodoropoulos), R.Bahsoon@cs.bham.ac.uk (Rami Bahsoon)

 Selection and/or peer-review under responsibility of Prof. Mitsuhisa Sato and Prof. Satoshi Matsuoka
doi:10.1016/j.procs.2011.04.190

The collusion problem is as a result of this reliance on members for recommendations. That is, each node keeps a record about the behaviour of other nodes of interest to determine their reputation [5, 6].

Consider a P2P network situation where two nodes A and B have been compromised by an adversary. If A and B exchange all of their secrets, then B can masquerade as A to all of B's neighbours that node A shares pair-wise keys with and vice versa. The keys subsequently obtained from other nodes can be reused by the attacker-controlled nodes (A and B), cascading the impact of the compromise. Therefore, an attacker can control a node undetectably by physically compromising the node, and the same node can in turn compromise other nodes [7]. Without counter-measures, the effects of this attack have been shown to dramatically affect both the security and network performance at run time as evidenced in poor reliability and poor quality of service, higher overhead and throughput degradation [2].

In previous papers [8, 9], we described how trust decisions could be corrupted through recommendations made by members. We proposed a framework that is capable of providing dynamic trust ratings of nodes at runtime and predicting the future behaviour of nodes through the simulation of historical and online behaviour. The framework does not rely on collective opinion and ratings to determine the reputation of system entities as it has been shown that such an approach can result in attacks such as collusion [7]. Instead, the framework predicts a potential compromise before the attack occurs so that informed decisions can be made, which may include isolating malicious members or even denying them service. The framework presented in the papers constituted a first step to exploit the Dynamic Data-Driven Application Systems (DDDAS) paradigm to aid the prediction of trust values.

Although it is assumed that the predictive power of RTMs depend on the supposition that past behaviour is an indication of future behaviour, this assumption might not be true with *intoxication attacks* [10, 11]. In this attack, a member behaves as expected for a sustained period of time to obtain a good reputation and only starts to misbehave afterwards. Intoxication makes it difficult for the system to identify such misbehaving members because of their high reputation. This attack occurs because the effect of past good behaviour outweighs the effect of current actions on reputation. This paper extends this supposition further by not only considering past behaviour, but also the possible future behaviour. Emphasis in the extended framework (presented in this paper) is placed on past histories, recent behaviour and the future behaviour of members.

Behavioural expectation in any context can be motivated from the social perspective, where individuals within a society are expected to behave in certain ways. A disreputable person could redeem himself through honest actions and a trusted person could become less reputable if they demonstrate deceit over time [12]. This implies that reputation and trust change over time, and are therefore dynamic. As reputation is a measure of trustworthiness, both trust and reputation will used synonymously in this paper.

In literature, trust and reputation are considered as static, we however argue that they are dynamic and call for an equally dynamic approach to computing and predicting trust. Dynamics of reputation is also reflected by its timeliness; reputation is aggregated over time by taking into account recent behaviour and past histories [13]. Time is a necessary dimension for reputation and this is the reason why the framework is useful; because it considers a node in different future scenarios to provide a more holistic view of domain events. Therefore, our approach is not only dynamic in making predictions that provide information about the system but also in good time.

In order for any RTM to fulfil its functions, observations and experiences that determine the reputation of each member have to be captured and represented numerically. That is, the qualitative information captured is converted to a quantitative one, referred to as ratings or *Trust Values* (TVs). The framework adopts a data-driven approach by using the information based on current behaviour, past histories and possible future scenarios in predicting an overall TV of members. Data from the reality is continually injected into the simulation at specified time intervals to aid the prediction. We argue that a more precise overall TV is obtained by considering *historical*, *online* and *future* (predicted) behaviour. This has been verified with experiments carried out. By comparing the results obtained with and without the use of prediction, the framework has shown to be more useful in terms of enabling informed decisions in time before misbehaviour occurs.

This paper is organised as follows: Section 2 describes existing reputation and trust-based models and Section 3 describes the dynamic framework. Some experimental analysis and results are presented in Section 4 while Section 5 contains discussion and future work.

2. Related work

Relevant models that have contributed significantly to reputation and trust management research are discussed in this section. The models have been suggested for use in diverse domains ranging from online communities to networks.

A significant reputation system for Mobile Ad-hoc Networks (MANETs) is that proposed by Buchegger *et al.* [2]. The RTM aims to detect and isolate misbehaving nodes by making it unattractive for any node to deny cooperation with others. In this protocol, each node maintains a reputation rating and a trust rating about every other node of interest. Nodes monitor and detect misbehaviour in their neighbourhood by means of an enhanced *packet acknowledgment*(PACK) mechanism where the confirmation of acknowledgment comes indirectly by overhearing the next node forward the packet.

EigenTrust [14], a P2P reputation system's algorithm, similar to the popular PageRank, works adequately if there are no colluding members. The aim of EigenTrust is to reduce inauthentic files distributed by malicious peers. The algorithm assigns each peer a unique global TV, based on the peer's history of uploads. EigenTrust's susceptibility to collusion has been demonstrated in [6], where certain colluding peers are able to obtain high trust values.

In the work of Ganeriwal *et al.* [3], applicable to wireless sensor networks (WSNs), each sensor node maintains reputation metrics. These metrics represent the past behaviour of other nodes and are used as an inherent aspect in predicting their future behaviour. The model relies on network members to maintain the reputation of others based on their experiences and uses this to evaluate their trustworthiness.

The online marketplace, eBay [15] has a reputation management system that uses a centralised approach for collecting and computing the ratings of users. The system has a feedback forum scheme where buyers and sellers rate each other for transactions based on their experiences with the other party. The system generally motivates buyers and sellers to be honest. Incidents such as buyers perpetuating intoxication attacks by building a high rating with low-valued transactions and then misbehaving with a high-valued one, have shown that there are drawbacks in the eBay system.

Current research on using DDDAS focus on the simulation of physical, artificial or social entities. The simulation is able to make predictions about how the entity would evolve and its future state. The predictions made by the simulation can then influence how and where future data will be gathered from the entity, in order to focus on areas of uncertainty [16]. This capability of the paradigm is exploited in this research to in data collection, simulation and prediction of TVs.

3. Data-driven framework

This section describes the dynamic framework that uses a data-driven approach to function. The sequence of events that take place for prediction and how the framework is useful in terms of timely feedback on future events is discussed in this section. The components of the framework are the physical system, simulation and the controller that computes trust. The components are depicted in Figure 1 while Figure 2 shows the steps required for trust computation, prediction and countermeasures in the framework.

3.1. Physical system

The physical system is divided into logical regions of high-risk, medium-risk and low-risk. The idea behind the division is that members are assigned to regions depending on their TVs. A member with a bad reputation and low TV belongs to the high-risk region and a member with a good reputation and high TV will belong to the low-risk region of the system.

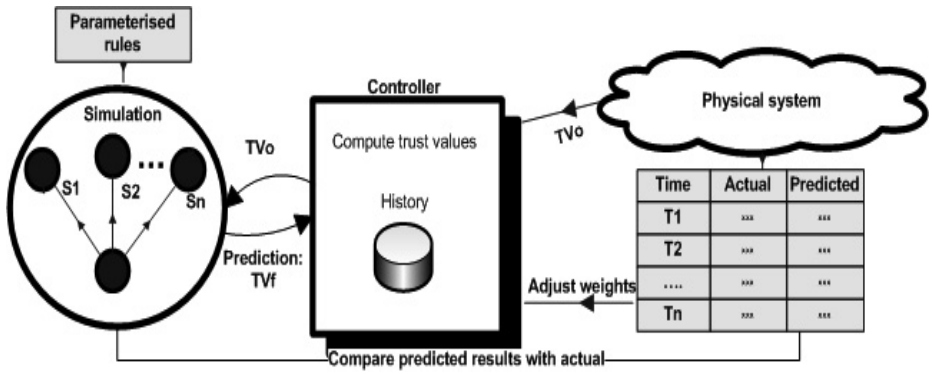

Figure 1: Framework components showing how data is injected into the simulation and the scenarios S1,S2,...Sn that are considered. tv_o and tv_f are the online and future trust values for every member.

The grouping into regions helps in the management of the network to focus on the critical group of members that require more attention. This will equally aid future informed decisions in the system.

3.2. Data collection

The simulation of the physical system runs concurrently with the system itself. However, the simulation is ahead in time of the physical system. At the start and at specified intervals, a picture of the system is captured and adapted in the simulation. This picture includes the current state of the system at a certain point in time.

For example, the TV of a certain node and all the connections to and from the node are captured. The TVs of connected nodes are also obtained from the system. This TV is the online TV (tv_o) that represents the behaviour of the nodes based on current domain events and the new TV (tv^R) that is computed from information about past histories and the online ratings.

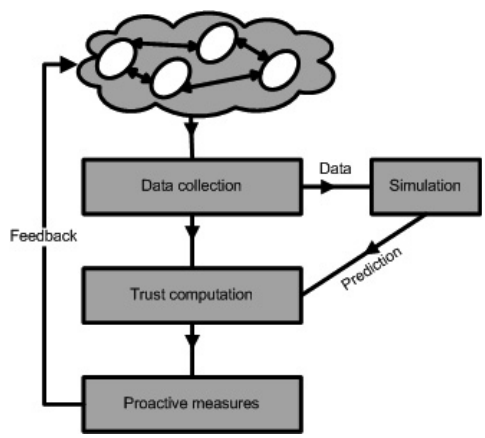

Figure 2: Trust prediction steps in the framework.

A set of discrete TVs are assumed in the framework and each value represents a degree of trust [8]. These discrete degrees of trust introduce flexibility into the application of our framework as different behaviour corresponds to different levels of trust.

Qualitative data captured is converted to a quantitative one by an associated TV. Collected data from the system is transformed to a value ranging from $[0, 5]$, where a score of 0 means a node is completely untrusted, 5 means a node is absolutely trusted and if $0 < TV < 5$, then it implies that the node is trusted to a certain extent.

3.3. Simulation

The values tv_o and tv_n from the physical system are injected into the simulation at the start. Using an agent based simulation tool, the simulation runs for more time steps and considers what-if scenarios which are possible states a member may be in the next time step or in the future. The number and type of scenarios will depend on the application domain. The state of a member depends on the behaviour exhibited by the member. Examples of the scenarios are: collusion attack; such as altering a message, intoxication and normal expected behaviour. After some specified time intervals $T_1, T_2, ..., T_n$, the simulation state is observed and compared with reality; which must have also evolved (see Figure 1). The framework is adaptive and if there are any differences in the predicted values and reality, the weights for the trust computation are continually adjusted to reflect reality.

Possible outcomes in the scenarios are simulated to anticipate possible fluctuations in member behaviour. This is because the behaviour of members generally in any network, domain or context is dynamic and changes over time. The scenarios are considered ahead of the time of the physical system. The TV is computed for a member in each scenario considered. With this information, it is possible to compute and anticipate the future TV of the member as described in the definitions below. In the controller, this information is combined with online and historical TVs to obtain an overall TV.

The following are sample rules that change the state of members in simulation. Rules can be added, removed, adjusted or customised depending on the domain of application. The rules below represent member behaviour in the context of a P2P network of file transfer between peers.

Rule 1: A member will have a neutral state at the onset. This implies that the node has a neutral trust value of 2.5.

Rule 2: A member's trust value will be decremented if it exhibits behaviour leading to collusion.

Rule 3: An member's trust value will be incremented if it actively participates and consistently transfer files according to the expected behaviour.

Rule 4: A member will be in a high risk region if its computed trust value is 1 and below. This implies that the state of the node is not trusted.

Rule 5: A member will be in a medium risk region if its computed trust value is above 1 but below 4.

Rule 6: A member will be in a low risk region if its computed trust value is above 4. This implies that the node is at a trusted state.

3.4. Trust computation

Computing trust in RTMs has been described as an abstract mathematical specification of how available information should be transformed into a usable metric [1]. This is also referred to as aggregation in some literature and it is a critical component that determines the reliability of the framework. Trust computation is difficult because is crucial to the fulfilment of the functions of any trust-dependent system and has to be defined in a precise way.

In this framework, the specification is made through explicit equations discussed in [8]. The online and historical trust values are obtained from the network and are used for the computation of the overall TV. The initial computation of tv^R is done without any consideration of any predicted values. The predicted values from the simulation are later considered in computing the overall value tv.

Definition 1: Using the notation tv^R to represent the initial TV in the physical system,

$$tv^R = \mu_h tv_h^R + \mu_o tv_o^R \tag{1}$$

This value only considers online data and past histories of each member. Weights μ_o, μ_h and μ_f are factors of online, historical and future TVs respectively and they are introduced to allow flexibility in the framework.

Definition 2: The simulation considers the possible scenarios a member may undertake in the future. The average of the ratings derived from these scenarios determines the future TV tv_f^S.

$$tv_f^S = \frac{(tv_o^S)S_1 + (tv_o^S)S_2 + \ldots + (tv_o^S)S_n}{N} \tag{2}$$

Therefore, an overall TV is computed by

$$tv = \mu_h tv_h^R + \mu_o tv_o^R + \mu_f tv_f^S \tag{3}$$

where tv_h^R, tv_o^R, tv_o^S and tv_f^S represent the historical, online TVs in the physical system, online TV in simulation and the predicted TV in simulation respectively. The number of scenarios considered in simulation is represented with N while $(S_1, S_2, \ldots S_n)$ are the scenarios.

The weights are used to control the effect of historical behaviour on the newly computed TVs. For example, if (μ_o, μ_h) > 0 and $\mu_o > \mu_h$, more emphasis is placed on recent behaviour as opposed to historical behaviour. The emphasis on recent behaviour prevents members from gaining a good reputation by behaving as expected over a sustained period of time and only then starts to misbehave (this is a form of intoxication attack as described earlier).

The framework is adaptable to changing domain conditions because the behaviour of members is compared with the predictions, if there are any inconsistencies; the weights are altered to reflect reality.

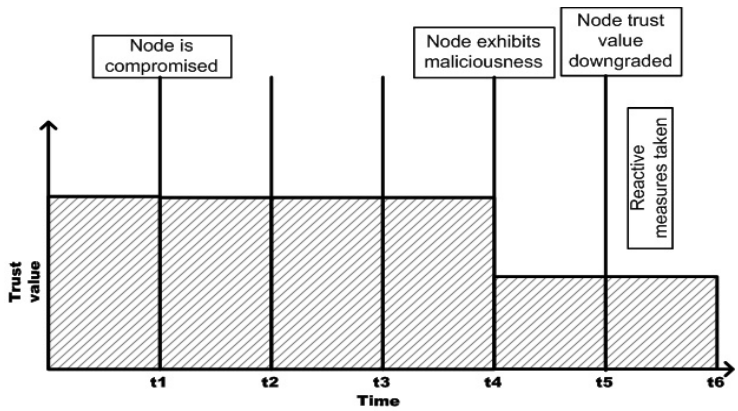

Figure 3: Other approaches

The framework performs better by predicting the future ratings of members. The prediction gives the system enough time for preventive measures, making the framework proactive compared to other models that are reactive in nature. The framework is proactive in terms of providing control such as downgrading the TV of suspect members that are misbehaving before they can carry out an attack. This is contrary to how other approaches work, which only downgrade the TV as a reaction to misbehaviour.

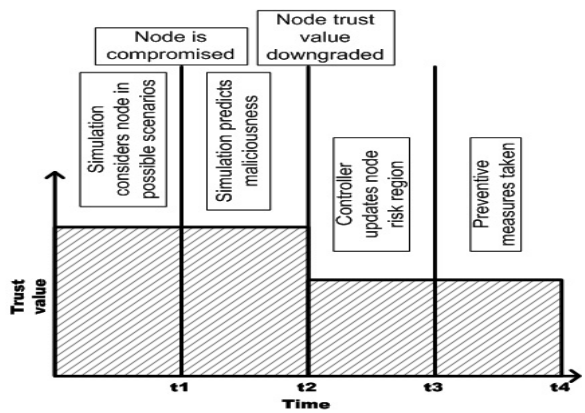

Figure 4: Dynamic approach

The assumption is that a member that has been compromised by an adversary for example, exhibits a sequence of behaviour in order to misbehave. The simulation component of the framework applies the past histories and current online behaviour and considers the compromised member in different scenarios to make predictions and ultimately, enable the framework to proactively control the system by downgrading TVs.

An example is depicted in Figures 3 and 4 which show the time delay between the framework and other approaches in predicting TVs and taking preventive measures. Figure 3 shows that the TV is only downgraded at time t_5 after the member exhibits maliciousness. The simulation in the framework predicts the maliciousness between time interval t_1 and t_2 and the TV is downgraded between t_2 and t_3 in Figure 4.

4. Analysis and results

In this section we present some experimental analysis to confirm the hypothesis described in Figures 3 and 4 showing the reliability of the framework in providing timely predictions.

4.1. Case study

Incentive systems in P2P networks have generally been vulnerable to peer collusion. Collusion patterns have been found in the Maze file-sharing system; a hybrid P2P network with central control described in [6]. In the paper, collusion is defined as a collaborative activity of a group of peers that grants its members benefits they would not be able to gain as individuals. Also considering the results of the research, the assumption in this framework is that collusion is in pairs because collusion between groups of three or more peers is rare.

Experiments are carried out on a P2P network scenario where the dynamic framework predicts the TV of network peers. Using a simplified dataset from Maze [17], let us consider a file sharing network between 100 peers with 4 colluding peers and 27 file transfers. In this experiment, files are shared in terms of messages transferred between peers.

TV ranges from 0 to 5 and a TV of 0 means no trust and that of 5 implies absolute trust. We take a cynical approach, where peers are not trusted on joining to the network. Each nodes has a neutral rating of 2.5 until they are able to demonstrate their trustworthiness or maliciousness.

The network is modelled with the following properties:

- Peers interact with other peers by considering the communication mechanism found in a P2P network, causing node states to change.
- The peers are self-contained as they are uniquely identifiable with a set of characteristics, behaviours and attributes.
- The peers function independently and interact with other peers by transferring files.
- The peers are situated in a physical network which is a grid with contexts that hold the peers

4.2. Implementation environment

This section describes the setup of the simulation environment. Our experiments are performed using an agent based modelling and simulation toolkit: Repast [18], with a mixture of Groovy and Java. The simulation parameters are shown in table 1. The network is constructed as a graph of peers that are represented as vertices. The edges between the vertices are the files transferred (one-way message exchange) between peers. The rules of interaction used in this experiment are those stated in Section 3.

The experiment is carried out with and without the predictive capability of the framework. What-if scenarios of collusion, intoxication and failure to forward files are considered in the simulation, as well as when the node behaves as expected.

4.3. Preliminary results

Two experiments were carried out to observe the effect of prediction on trust computation. The first experiment was with prediction and the other was without prediction. Figure 5 shows the results of using and not using prediction in the same P2P network scenario. The figure shows the trust value of one of the malicious nodes over time. With prediction, the framework detected and flagged the 4 peers as malicious at 40 ticks and their ratings were downgraded immediately, to anticipate the future.

Table 1: Simulation parameters

Parameter	Value
Total simulation time (in ticks)	100
Total number of nodes	100
Number of malicious nodes	4
Total number of messages transferred	27
Default historical trust value tv_h	2.5
Default online trust value tv_o	2.5
Online weight μ_o	0.5
Historical weight μ_h	0.3
Prediction weight μ_f	0.2

This result is different from the second experiment where the same data set was used but without prediction. In the second experiment, the downgrade of the trust value began at 60 and reflected peer reputation after the peer had carried out the attack.

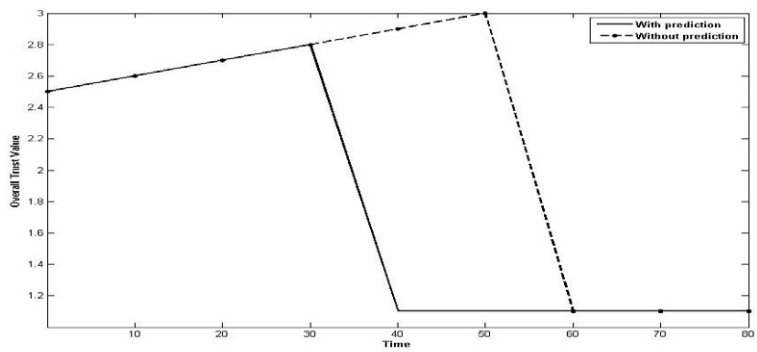

Figure 5: P2P file-sharing network result (with and without prediction of trust values)

The trust values of the malicious peers dropped below the threshold value of 2. Therefore, the peers now logically belong to the high-risk region of the physical system. Ultimately, the nodes are isolated because their TVs were below the threshold for other peers to want to cooperate with them. Figure 6 compares the predicted trust with actual TV for some peers. The graph shows the changes in the value of a peer exhibiting intoxication, a non-participating peer that does not cooperate in terms of transferring files and thus, maintains a stable value, an untrusted peer whose TV continues to drop, and a trusted peer that is active with a high value.

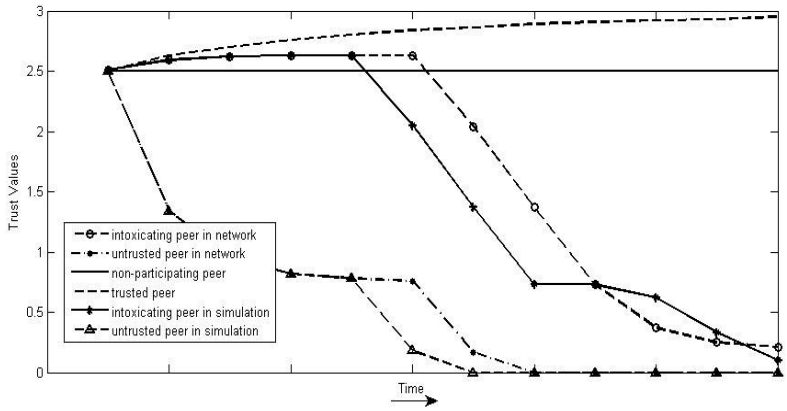

Figure 6: Trust values of peer and the comparison of the values in the network and simulation

A comparison of the simulation results with the real physical system indicated some degree of variance between the actual and predicted trust values. This might account for possible false-positives or false-negatives generated from our simulation. Hence, we shall explore approaches to improve the correlation of these trust values (i.e. simulated and real system) in the future.

5. Discussion and future work

This paper presents a novel framework that predicts ratings and flags malicious members in good time for security mitigation in the system. This can potentially improve the reliability of the system. The framework was applied with and without prediction to the same dataset and the observation is that the framework is more proactive with predictive capability. The approach adopted in the framework treats the problem of collusion attack by not relying on recommendations made by members, since such members can actually corrupt the system.

When compared to a distributed reputation and trust-based model, a centralised predictive framework may have a relatively higher performance overhead, but it is more timely in terms of misbehaviour detection as illustrated in Figures 3 and 4. Thus, we propose the use of our approach in niche environments where the security requirement is more critical compared to other non-functional requirements.

The success of a trust-based model is measured by how accurately the computed values reflect the reality of future interactions among members. Therefore, in the future, we aim to test the accuracy of predictions in reflecting the reality. We will also consider approaches to parameterise the rules that change the state of members for more dynamism in the framework in the near future.

References

[1] K. Hoffman, D. Zage, C. Nita-Rotaru, A survey of attack and defense techniques for reputation systems, ACM Computing Surveys 42 (1) (2009) 1–31.

[2] S. Buchegger, J. Le Boudec, Performance analysis of the CONFIDANT protocol (Cooperation of nodes: Fairness In Dynamic Ad-hoc Networks), in: Proceedings of the International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc, 2002, pp. 226–236.

[3] S. Ganeriwal, L. K. Balzano, M. B. Srivastava, Reputation-based framework for high integrity sensor networks, ACM Transactions on Sensor Networks 4 (3) (2008) 15:1–37.

[4] P. Michiardi, R. Molva, CORE: A collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks, in: Proceedings of the IFIP TC6/TC11 Sixth Joint Working Conference on Communications and Multimedia Security, Vol. 100, Kluwer, B.V., 2002, pp. 107–121.

[5] J. Hu, M. Burmester, LARS - A locally aware reputation system for mobile ad hoc networks, in: Proceedings of the ACM SE Regional Conference, Vol. 2006, 2006, pp. 119 – 123.

[6] Q. Lian, Z. Zhang, M. Yang, B. Y. Zhao, Y. Dai, X. Li, An Empirical Study of Collusion Behavior in the Maze P2P File-Sharing System, Proceedings of the 27th IEEE International Conference on Distributed Computing Systems 0 (2007) 56.

[7] T. Moore, A collusion attack on pairwise key predistribution schemes for distributed sensor networks, in: Proceedings of the 4th Annual IEEE International Conference on Pervasive Computing and Communications Workshops, Vol. 2006, 2006, pp. 251 – 255.

[8] O. Onolaja, R. Bahsoon, G. Theodoropoulos, Conceptual framework for dynamic trust monitoring and prediction, Procedia Computer Science 1 (1) (2010) 1235 – 1244, ICCS 2010.
[9] O. Onolaja, R. Bahsoon, G. Theodoropoulos, An Architecture for Dynamic Trust Monitoring in Mobile Networks, in: OTM 2009 workshops, Vol. 5872 of Lecture Notes in Computer Science, 2009, pp. 494–503.
[10] P. Kollock, The Production of Trust in Online Markets, in: Advances in Group Processes, Vol. 16, 1999, pp. 99 – 123.
[11] S. Buchegger, J. Le Boudec, Self-policing mobile ad hoc networks by reputation systems, IEEE Communications Magazine 43 (7) (2005) 101–107.
[12] W. Adams, I. Davis, N.J., Toward a decentralized trust-based access control system for dynamic collaboration, in: Proceedings from the Sixth Annual IEEE Systems, Man and Cybernetics (SMC) Information Assurance Workshop, 2005, pp. 317 – 324.
[13] J. Liu, V. Issarny, Enhanced reputation mechanism for mobile ad hoc networks, in: Trust Management. Second International Conference, iTrust 2004. Proceedings.(Lecture Notes in Comput. Sci. Vol.2995), 2004, pp. 48 – 62.
[14] S. Kamvar, M. Schlosser, H. Garcia-Molina, The Eigentrust algorithm for reputation management in P2P networks, in: Proceedings of the 12th international conference on World Wide Web, WWW '03, ACM, 2003, pp. 640–651.
[15] eBay, Online, `http://www.eBay.co.uk` (accessed 21 Jan 2011).
[16] C. Kennedy, G. Theodoropoulos, Intelligent management of data driven simulations to support model building in the social sciences, Computational Science-ICCS 2006. 6th International Conference. Proceedings, Part III (Lecture Notes in Computer Science) 3993 (2006) 562 – 569.
[17] M. Yang, H. Chen, B. Y. Zhao, Y. Daiand, Z. Zhang, Deployment of a Large-scale Peer-to-Peer Social Network, in: Usenix workshop on real, large distributed systems (WORLDS), 2004.
[18] Repast, Online, `http://repast.sourceforge.net` (accessed 21 Jan 2011).