Published in the Proceedings of the 2007 IEEE/WIC/ACM International Conference on Web Intelligence (WI'07)
Silicon Valley, Nov 2007, pp 676-682.

A Multi-scenario Reputation Estimation Framework and its Resilience Study against Various forms of Attacks

Javed I. Khan
Media Communications and Networking Research Laboratory
Department of Computer Science,
Kent State University, Kent, OH 44242
javed@kent.edu

Sajid S. Shaikh
Brulant Inc
3700 Park East Drive
Beachwood, OH 44122
sajid.shaikh@brulant.com

Abstract

Online transactional activities that involve establishment of trust between participating individual seem to require a reputation function for the reputation estimation framework (REF). They are often vulnerable to various kinds of attacks. Also it seems we do not evaluate reputation in the same way in all situations. Using Occam's razor we propose a generalized set-theoretic reputation function with customizable components that can be changed to meet the reputation requirements in wide variety of reputation assessment scenarios. Further we identify several canonical classes of the functions. The resilience of the framework is then analyzed by subjecting it to various reputation attacks such as gang attacks, vendetta and Dr Jekyll & Mr. Hyde.

1. Introduction

Reputation is an essential component of social mechanics where to commit various social transactions one party must need to infer probable response of the other while taking a risk. Reputation is defined as 'socially constructed labels that extend the consequences of a party's actions across time, situations, and other actions' [1]. Online reputation is also increasingly attaining significance. Lei has shown that sellers with good reputation have a higher probability of sale and get higher transaction price. Akerlof [2] has shown that a market breakdown is highly likely if the sellers cannot convince the buyers about the quality of goods and services. It seems that reputation in particularly important in the online world where perfect strangers are meeting in numbers which has no precedence. High reputation can help in instilling more confidence in the buyer about the seller. Battalio, Ellul & Jennings [3] have shown that reputation plays an important role in the liquidity provision process on the floor of the NYSE. Ghose et. al [4] suggest that a seller with better reputation can successfully charge higher prices than competing sellers of identical products, and that their pricing power increases with their recorded level of experience.

The Internet is also expanding the variety of interactions in which perfect strangers are engaging. This is manifested in the emergence of social networks surrounding numerous digital services. On the Internet there is a growing trend of like-minded people coming together to form virtual communities. The communities thus formed indulge in diverse activities ranging from simple friendly networking to ecommerce. The web is filled with online social networks, ecommerce websites and P2P systems. A lot of data and information is being shared among individuals. The building block for this kind of cooperative behavior is the mutual trust between various individuals. The trustworthiness of an individual is positively correlated to his/her reputation.

As community size grows, the individuals are exposed to scenarios where they have to interact with unknown individuals. In these scenarios, reputation seems to be a good scale to measure against before interacting with strangers. A higher individual reputation most likely correlates positively with individual trustworthiness. In spite of these precautions, one can be faced with attack scenarios where somebody is trying to compromise ones reputation. Nielson et al ([5]) have identified and created taxonomy for rational attacks and then identified the corresponding solutions if they exist. The threats enumerated by Dellarocas ([6]) are similar to the kind of threats we have tackled through our REF. Reputation seems to be a very important part of *social mechanics*. It seems we as a social being employ some form of computation to assess reputation of people in numerous situations. An interesting question is to see what might be the functional and computational form

Published in the Proceedings of the 2007 IEEE/WIC/ACM International Conference on Web Intelligence (WI'07)
Silicon Valley, Nov 2007, pp 676-682.

of it? Do we all use similar function? Do we use different functions in different scenarios? Is their any resilience against various attacks? Currently, we could find few very simple adhoc functions.

In this research we take a holistic view to this problem and look closely at various factors which generally is attached in the assessment of reputation. We then suggest a generic reputation function for quantifying the reputation of a peer in any community-like environment. Our goal is to have an attack tolerant generic system, which is dynamic and customizable.

In this paper, we present this new generic function in section-2. It can be used inside a reputation management system (RMS) framework. Section-3 presents a discussion on various threats we generally encounter over such framework. Sections-4 presents the resilience characteristics of the suggested function under several forms of attacks.

2. Reputation Model

In this section, we present a social-transactional model of a generalized reputation estimation system framework. This is followed by a discussion of the various factors that influencing the reputation of a peer and towards the end we present a mathematical formulation for quantifying reputation.

Reputation is the estimation of an individuals status in a social setup. It is built based on various social transactions and the evaluations of those transactions. Thus, these transactions collectively builds up a memory about a target individual and this is estimated in target's reputation function. The value is useful to establish trust in a later transaction involving the target in various communities.

Social science literature identifies a number of factors that affect reputation, but we consider the ones, that are the most dominant.(i) Opinion about a Transaction (O): Generally, each transaction creates an evaluation about the goodness of a peer. Reputation relies on these individual feedbacks or opinions to evaluate a stable measure about the goodness of a peer. (ii) Reputation of Opinion Provider (R): Whenever a peer expresses an opinion, many social scenarios seem to take into account as to who exactly is providing this opinion. This factor helps in weighing certain opinion more heavily than others weigh. (iii) Age of the Opinion (T): This factor captures the freshness of the opinions. The opinion gets older with time and hence its impact on the overall reputation becomes less. (iv) Number of Transactions (N): An averaging function is a better mathematical representation to determine reputation. Hence, the total number of transactions is an important factor in determining reputation irrespective of the volume of transaction. (v) Group Reputation (W): Group reputation is taken into consideration to somewhat depict the real society where individuals with high reputation tend to be associated with a group whose members are also highly reputed. Nevertheless, in cases where they are surrounded with low reputation, they encourage the lowly reputed members to improve their respective reputation. (vi) Impact Parameters (X & α): These variables are used to control the direction of influence and the amount of influence the above-mentioned variables would have on the overall reputation of the peer. The above parameters can be functionally connected in many different forms. We use Occam' razor principle and suggest the following generic reputation function (equation 1).

$$R_A(t) = \sum_{k=1}^{m} W_k \left[\frac{\sum_{j=1}^{N} R_j^{\alpha^{R\times X^R}} \times O_j^{\alpha^{O\times X^O}} \times e^{(-\lambda Tj)\alpha^{T\times X^T}}}{N^{\alpha^{N\times X^N}} + \sum_{j=1}^{m} W^{\alpha^{W\times X^W}}} \right] + \Phi e^{-\lambda / T_n} \quad (1)$$

2.1 The Generic Reputation Function

Reputation in a society seems to be positively correlated to opinion, individual reputation of opinion provider and opinion freshness. Hence the generic reputation function is a product of the three variables We have used an averaging function instead of a summation function since we wanted to restrict the value between 0 and 1. The decrease in the freshness of opinion is a gradual process, hence we have used an exponential function. Each factor can affect the reputation evaluation process either positively, negatively or have no impact (zero impact).The impact variable X controls the influence direction of the various factors. Each factor has its one independent impact variable. Certain factors may be more aggressively involved in the evaluation process as compared to others. This behavior can be captured by the impact weight variable α. Each factor has its own controlling α and by assigning the appropriate values, the impact of certain variables can be made more pronounced as oppose to others.

2.2 Canonical Classes of the Function

Depending upon the deployment environment, certain variables would impact the reputation where as others wont' be part of the determination process. There are four primary customizable variables viz. R, T, N and W, thus there are sixteen possible ways to customize them. We could find at least five of these combinations have corresponding real life examples. We call then canonical reputation scenarios. Table 1 lists these canonical scenarios.

Name	Function	Example
Fading Memory Averaging Function	$R_A(t)=\left[\frac{\sum_{j=1}^{N} R_j^{\alpha^{R\times X^R}} \times O_j^{\alpha^{O\times X^O}} \times e^{(-\lambda Tj)\alpha^{T\times X^T}} \times W^{\alpha^{W\times X^W}}}{\sum_{j=1}^{N} e^{(-\lambda Tj)}}\right]+\Phi e^{-\lambda/T_n}$	Readers expressing opinions about a book.
Memory less Summation Function	$R_A(t)=\sum_{j=1}^{N} R_j^{\alpha^{R\times X^R}} \times O_j^{\alpha^{O\times X^O}} \times W^{\alpha^{W\times X^W}}$	Authors expressing opinion about their book
Fading Memory Averaging Function without Opinion Credibility	$R_A(t)=\left[\frac{\sum_{j=1}^{N} R_j^{\alpha^{R\times 0}} \times O_j^{\alpha^{O\times X^O}} \times e^{(-\lambda Tj)\alpha^{T\times X^T}} \times W^{\alpha^{W\times X^W}}}{\sum_{j=1}^{N} e^{(-\lambda Tj)}}\right]+\Phi e^{-\lambda/T_n}$	Viewers expressing opinion about a movie.
Fading Memory Averaging Function without Community Context Factor	$R_A(t)=\left[\frac{\sum_{j=1}^{N} R_j^{\alpha^{R\times X^R}} \times O_j^{\alpha^{O\times X^O}} \times e^{(-\lambda Tj)\alpha^{T\times X^T}} \times W^{\alpha^{W\times 0}}}{\sum_{j=1}^{N} e^{(-\lambda Tj)}}\right]+\Phi e^{-\lambda/T_n}$	Critics providing opinion about a movie.
Memory less Averaging Function	$R_A(t)=\left[\frac{\sum_{j=1}^{N} R_j^{\alpha^{R\times X^R}} \times O_j^{\alpha^{O\times X^O}} \times W^{\alpha^{W\times 0}}}{N}\right]$	

Table 1: Canonical classification of the reputation function

3. Threats to the Model

In the next sections, we have presented a detailed description of the various attacks that are possible on reputation management systems. In section 3.1, we introduce the various parties involved in the attack followed by an explanation of the various attacks in section 3.2

3.1. Parties Involved in an Attack

(i) Attacker(s) (AT): The attacker/perpetrator can either be the person giving an opinion about the target individual or the target individual himself. We assume that the attacker always lies. The attacker person/group can be of three types. (1) Average Group: It contains a mixture of members having high reputation and low reputation. (2) Very Good Group: All the members of this group have high reputation. (3)Very Bad Group: All the members of this group have low reputation. (ii) Evaluators (EV): They represent the general population and are essentially the Controller Conglomeration, which provides random correct opinion about the target individual. We assume in our system that the evaluators never lie and are always truthful. The Evaluator contains members who have a range of reputations from high to low. (iii) Target (TG): The target could be a single individual of a group of individuals. (iv) Offender (OF): The offender is the person who commits something bad in the system for which he should be penalized.

3.2. Various Reputation Attacks

3.2.1. Vendetta

An attacker may target a single user by giving him a low opinion. This attacker could have High, Low or Average Individual Reputation. The impact of the attack differs depending upon the attacker's individual reputation.

3.2.2. Damaging Gang Attack

The attacker can join a group of other attackers to lessen the reputation of the target. The attacking group

provides unfairly negative opinions to the targeted good user, thereby lowering his reputation.

3.2.3 Dr Jekyll & Mr. Hyde

An offender starts off in the system in a well-behaved manner. As a result, his reputation in the system goes up. Once his reputation is sufficiently high he turns hostile.

4. Experimental Evaluation

We performed three sets of experiments to evaluate our reputation model. Through these experiments, we prove that out model stands it ground in the face of different attacks. The graphs are plotted with final reputation on Y-axis versus the time of the opinion on the X-axis.

4.1 Vendetta

The evaluator population is random providing honest opinion to the target. We have a single target and a single attacker.

4.1.1 Class One Function Vendetta Results

The two ellipses in figure 1 denote different periods of attack. Overall from figure 1, we can deduce that personal attack has a very limited or no damaging effect on the target reputation if the attacker frequency is low but can have a considerable impact in case of higher attacker frequency.

4.1.2 Class Three Function Vendetta Result

The results for this function show a behavior similar to the one shown by the fading memory averaging function.

4.1.3 Class Four Function Vendetta Results

The results for this function show a behavior similar to the one shown by the fading memory averaging function.

4.2 Damaging Gang Attack

We have a single target and a group of attackers. The attackers are initially part of the evaluator group but abruptly turn hostile. The number of members of the attacker group is set to 10 % of the total number of evaluator peer.

4.2.1 Class One Function Damaging Gang Results

In figure 4, we can notice small recoveries of the target's reputation; one of them is represented by the dotted circle. Through figure 4, we observe that though the attackers manage to bring down the reputation of the target during the attack period, they are not able to inflict permanent damage. The function recovers itself to the original value through the honest opinion expressed by evaluators with high reputation and the age of the opinion variable.

4.2.2 Class Three Function Results

We observe a behavior similar to the one discussed in section 4.2.1. The reputation goes down during the attack period but then the recovery starts as soon as the attack is over as depicted in figure 5.

4.2.3 Class Four Function Results

The function results shown in figure 8 are similar to the fading memory averaging function discussed in section 4.2.1.

4.3 Dr Jekyll & Mr. Hyde

The simulation consists of three groups. The evaluator group in this scenario consists of the average group, the very good group and the very bad group. The simulation design is graphically represented in figure 7.

4.3.1 Class One Function Results

The Dr Jekyll and Mr. Hyde phenomenon is vividly seen in figure 7. The evaluators punish the target for his offence, which results in his reputation taking a downward slide. However, he recovers his reputation through indulging in honest transaction, again to commit offence for which he is duly penalized. This trend is seen in figure 7 by the upward and downward movement of the reputation function.

4.3.2 Class Three Function Results

The results observed are similar to the fading memory averaging function results as shown in fig- 8.

4.3.3 Class Four Function Results

The results observed are similar to the fading memory averaging function as shown in figure 9.

Published in the Proceedings of the 2007 IEEE/WIC/ACM International Conference on Web Intelligence (WI'07)
Silicon Valley, Nov 2007, pp 676-682.

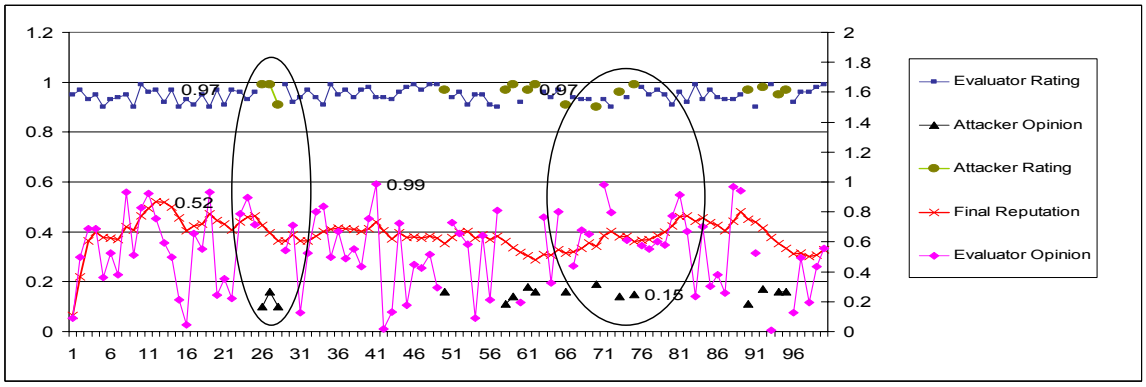

Figure 1: Class one function behavior when attacker has random personal reputation

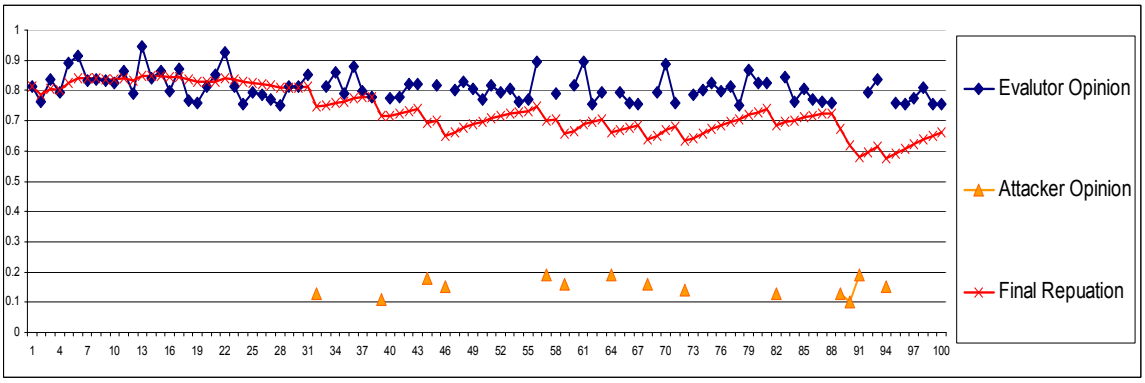

Figure 2: Class three function behavior during Vendetta.

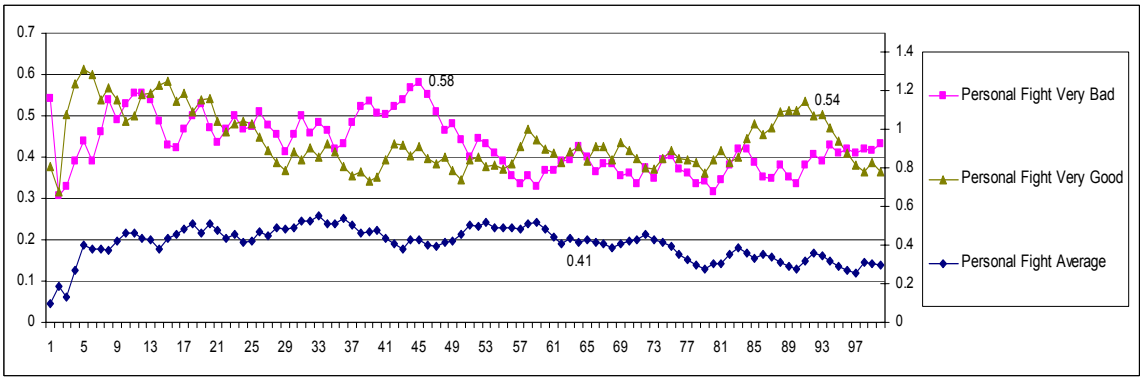

Figure 3: Class four function behavior during Vendetta

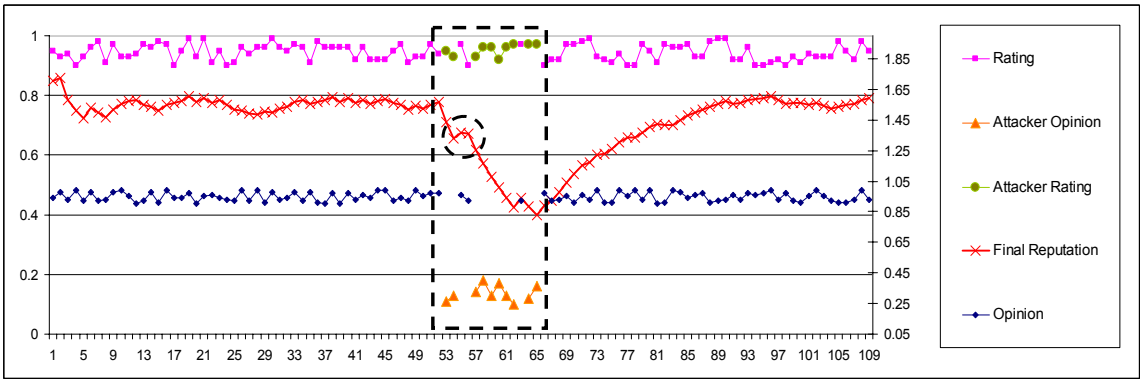

Figure 4: Class one function behavior when attacker group members have high personal reputation

Published in the Proceedings of the 2007 IEEE/WIC/ACM International Conference on Web Intelligence (WI'07)
Silicon Valley, Nov 2007, pp 676-682.

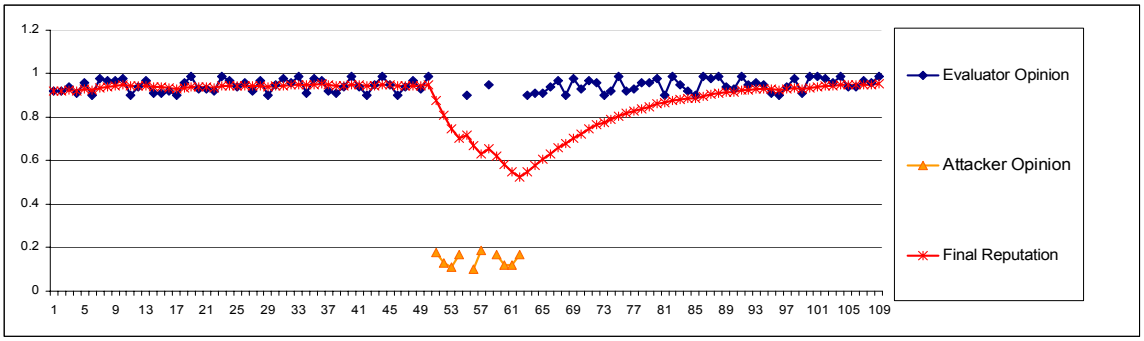

Figure 5: Class three function behavior under Damaging Gang Attack

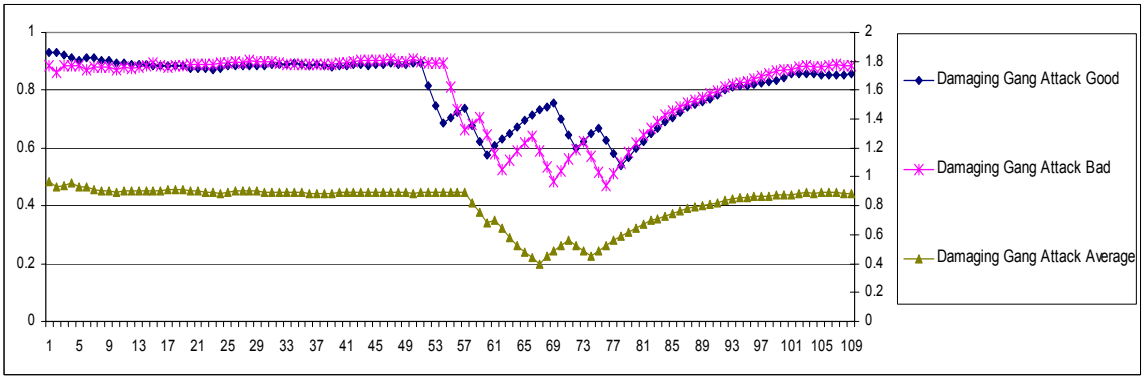

Figure 6: Class four function behavior under Damaging Gang Attacks

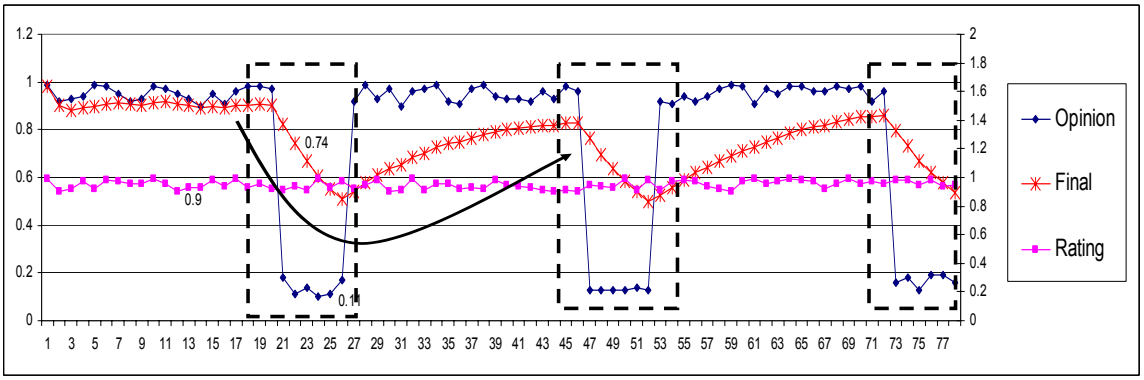

Figure 7: Class one function behavior evaluator group members have high personal reputations.

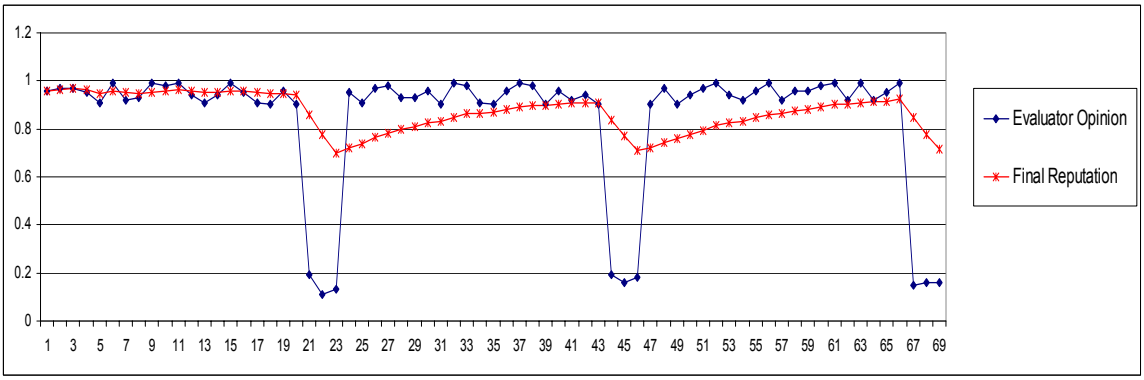

Figure 8: Class three function behavior during Dr Jekyll & Mr. Hyde attack

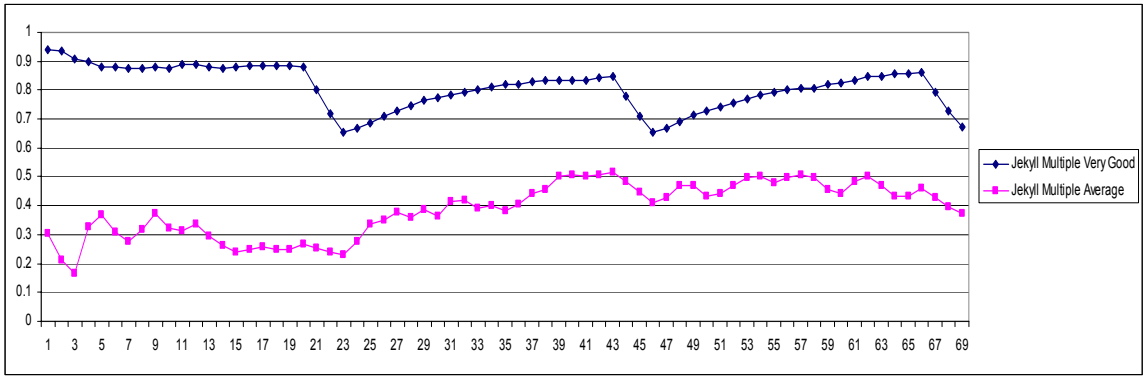

Figure 9: Class four function behavior for Dr. Jekyll & Mr. Hyde.

5. Resilience Strategies

Naturally, full attack tolerance can not be achieved just in one layer. It will require an integrated approach involving other components of the online system- particularly involving identity management, authentication, and non-repudiation processes of the overall system. A good reputation function should help detection. Through simulations we have shown the behavior of the functions under various attacks or *attack signatures*. The signature can provide important guidance towards the design of a resilient framework. Every consistent rapid change in reputation slope or reputation shift should be a suspect of possible attack. It is possible that a reputation shift is genuine. However, a suspect case should be subject to further analysis. Vendetta attacks can be detected and possibly resisted by looking at the temporal frequency and rating uniformity of the origination point of the opinions. Gang attack will require a more generalized approach. Restricting the rating frequency for evaluations which are showing strong spatial (distance in social network) and/or temporal locality can help thwarting distributed yet organized attacks. The slop characteristic can provide the required limits. However, some problems are even harder. For example there is no definite way of distinguishing between Damaging Gang and Dr. Jekyll & Mr. Hide kind of attacks. The number of low opinions being expressed towards the target can give some kind of indication as to which attack is in progress is 6.

6. Conclusions

In this paper we have discussed a probable functional form towards estimation of generalized reputation. To our knowledge this is one of the first such works. Definitely we as a social element employ some form of computation to assess and update reputation of people with whom we deal with. Reputation seems to be a very important part of our social mechanics. It seems online reputation will be increasingly high stack asset. In the corporate world there are too many examples where millions have been lost due to loss reputation. Large stake holders will aggressively guard their online reputations. Also attempt will intensify to compromise it. We envision some form of *detective units* outlined in section-5 to be become an integral part of any reputation framework of the future for continuous monitoring of suspected shifts and selected raters.

7. References

[1] C. H. Tinsley, K. M. O'Connor, & B. A. Sullivan. Tough guys finish last: The perils of a distributive reputation. Organizational Behavior & Human Decision Processes, 88(2), 621-642. 2002

[2] A.G. Akerlof, The Market for 'Lemons': Quality Uncertainty and the Market Mechanism, The Quarterly Journal of Economics, MIT Press, vol. 84(3), pages 488-500, August.1970.

[3] A. Ellul, R. H Jennings, and R. H. Battalio, Reputation Effects in Trading on the New York Stock Exchange. AFA 2006

[4] Ghose, Anindya, Ipeirotis, Panagiotis G. and Sundararajan, Arun, The Dimensions of Reputation in Electronic Markets (February 2006). NYU Center for Digital Economy Research Working Paper No. CeDER-06-02

[5] S. J. Nielson, S.A. Crosby, D.S. Wallach. A Taxonomy of Rational Attack.

[6] C. Dellarocas. Immunizing Online Reputation Reporting Systems Against Unfair Ratings and Discriminatory Behavior. ACM, Minneapolis, Minnesota, USA 2000.