TrustDavis: A Non-Exploitable Online Reputation System

Dimitri do B. DeFigueiredo
University of California, Davis
Department of Computer Science
Davis, USA
defigueiredo@ucdavis.edu

Earl T. Barr
University of California, Davis
Department of Computer Science
Davis, USA
barre@acm.org

Abstract

We present TrustDavis, an online reputation system that provides insurance against trade fraud by leveraging existing relationships between players, such as the ones present in existing social networks. Using TrustDavis and a simple strategy, an honest player can set an upper bound on the losses caused by any malicious collusion of players. In addition, TrustDavis incents participants to accurately rate each other, resists participant's pseudonym changes, and is inherently distributed.

1 Introduction

Some online auction sites have formalized the means by which individuals provide feedback on buyers and sellers. Loosely speaking, we call such mechanisms *online reputation systems*. "A reputation system collects, distributes, and aggregates feedback about participants' past behavior"[13]. Examples are eBay's Feedback Forum[1] and the feedback ratings at overstock.com.

Such systems usually attribute a rating to a particular identity. Ideally, individuals with good ratings are reliable trade partners, whereas individuals with poor ratings should be avoided[2]. Unfortunately, the reputation systems now available on the internet can be manipulated by malicious individuals or groups for selfish purposes. For example, a group can collude to artificially improve an individual's ratings with the intent of tricking unsuspecting victims into trading with someone that will never deliver the goods. This is the well known "hit and run" problem, to which all unsecured bilateral exchange is susceptible as there is always the temptation to receive a good or service without reciprocation [10]. This problem is aggravated online as many trade partners are veiled by relative anonymity and rarely trade. Mechanisms that have been proposed to mitigate such problems have achieved limited success [5, 1]. Ideally, we want a reputation system that resists malicious manipulation by groups of colluding parties, or at least that provides a strategy that honest participants can use to limit their exposure to such manipulation. TrustDavis has this property, as well as three others, these properties are:

- Honest participants can limit the damage caused by *malicious* collusions of dishonest participants.
- Malicious participants gain no significant advantage by changing or issuing themselves multiple identities.
- There is strong incentive for participants to provide *accurate* ratings of each other.
- It requires no centralized services, and thus can be easily distributed.

To our knowledge, TrustDavis is the first online reputation system proposed that can provide hard limits on the risk exposure of participants and combines these properties.

The outline of the paper is as follows. Section 2 briefly reviews of the current literature, focusing on motivating the three properties not yet discussed in detail. Section 3 describes the basic framework of the system, the use of references. Sections 3.1 and 3.2 obtain upper and lower bounds on the price of references. Section 4 describes a strategy that helps honest players avoid exploitation by malicious ones. In section 5, we summarize our results and provide suggestions for further research.

2 Related Work

An important difference between "real world" reputations and online reputations is that it may be possible to shed a bad online reputation by simply changing one's online pseudonym. This is a challenge that reputation systems should address [13]. This "cheap pseudonym" characteristic of online interaction imposes fundamental constraints on

[1] See the eBay website at www.ebay.com for a description of the Feedback Forum.

[2] Here we assume that past behavior is indicative of future behavior.

the degree of cooperation that can be achieved and to how well newcomers are treated by the community [6]. In TrustDavis, the ability to issue multiple identities or to change identity does not provide a significant advantage to a malicious party, since a malicious party must back each identity with funds that other players can use to protect their transactions.

We see *online reputation systems* and *trust inference protocols* as two sides of the same coin, since both mechanisms deal with the trust transitivity problem. In both cases, one must infer the reliability (trustworthiness) of an agent based on one's own experience and the experience of others. There have been quite a few proposals in the literature that address the trust transitivity problem each with its own set of desiderata [8, 7, 14, 16, 2].

Some proposals that address the trust transitivity problem allow each party arbitrary control over the ratings they provide [7]. Thus, one individual may rate all of his acquaintances as extremely reliable (trustworthy). This offers flexibility, but it may also allow malicious parties to trick the system by rating other parties undeservedly well. Some improvement on the "quality of the ratings" can be achieved if individuals are not allowed to rate others arbitrarily. A good example of such an approach is the EigenTrust trust inference algorithm for peer-to-peer networks [8]. In EigenTrust there is a normalization step that implies that peers only have a limited amount of trust to assign to each of their neighbors. In fact, one can argue that peers are only assigned a relative trust value in EigenTrust not an absolute one[3]. We believe it is desirable to ensure honest reporting of the past behavior of other participants as pointed out in [13, 12]. In the system we propose, there is a strong incentive for individuals to rate others accurately through references, since they are liable for bad references.

Usually, one of the goals of having a reputation system is to elicit better behavior from participants by providing the right incentives. It can be very useful in distributed systems such as peer-to-peer networks to make systems "incentive compatible". For example, the free-rider problem can be seen as incentive compatibility issue. Thus, it is desirable to have a reputation system that can be distributed to enable its deployment in distributed applications such as peer-to-peer networks. TrustDavis depends solely on paths between transacting parties and is as a result, inherently distributed.

3 The Model

We view agents or players as vertices in a graph $G = (V, E)$. Initially, agents publish information about other agents whom they know and trust, by publishing references to them. Each reference is an acceptance of limited liability. We expect existing social relationships to be represented after this initialization. Parents would give references to their kids and spouses. Business partners would give references to fellow workers, friends would provide references to friends and so on. Individuals with no references can join TrustDavis through the use of security deposits. They would simply leave a deposit with a member of the network that would then provide references to the newcomers. The newcomers should choose a trustworthy member for this task. This points out two issues:

- Parties assume liability (take risks) when they provide references and thus references should be provided only to trusted parties.
- There should be some incentive for parties to provide references and take on risk. Thus, parties can function as insurers and charge a premium for the references they provide.

If player A gives a reference to player B in the value of \$100, then player A would be willing to accept limited liability for bad trade caused by B. In other words, if B were to default payment on a transaction, A would be willing to pay the creditor up to \$100. Similarly, if B failed to ship a product, A should be willing to reimburse the buyer for payments already made up to the total of \$100. We say that A *would* be willing to accept liability because the reference is only a statement of A's intent. Before A accepts liability she needs to check two things:

- Whether someone else is already using the reference requested; and
- Who is asking for the reference.

This should be done in real-time (online) to avoid duplicate usage of a single reference.

In our graph G, there is a directed edge going from vertex v_1 to vertex v_2 if v_1 gives v_2 a reference. Each edge is labeled by the value of the reference provided and each edge label can be seen as the maximum flow capacity for that edge. Note that a vertex controls the "flow" on all its outbound edges by controlling the references it provides to other parties.

If vertex v_b wants to buy a product valued at \$$x$ dollars from vertex v_s then both vertices can complete the transaction with *no risk* if the aggregate *network flow capacity* from v_b to v_s and vice versa is of a value larger than or equal to x. To insure himself against bad behavior from vertex v_s, the buyer v_b obtains enough "flow capacity" to cover the value of the transaction from each vertex in the paths from v_b to v_s. Similarly, to insure the transaction with v_b, the seller v_s

[3] Unfortunately, it is still possible to trick EigenTrust into attributing an undeservedly high rating to an individual by obtaining multiple identities.

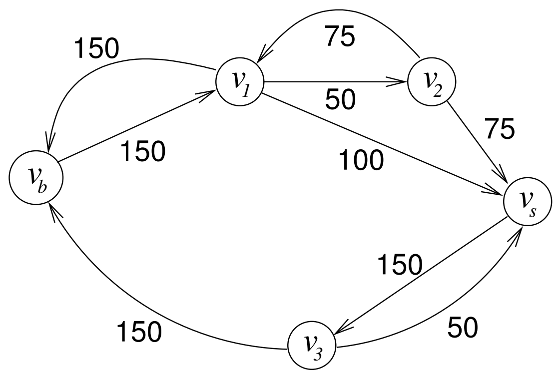

Figure 1. A simple example network, edges denote how much liability nodes are willing to accept.

also obtains references from enough vertices so that the aggregate flow in the opposite direction (*i.e.* from v_s to v_b) is at least equal to the value of the transaction. In a sense we are going to reduce the trust transitivity problem to a network flow problem.

Consider the situation depicted in Figure 1. Assume all vertices are willing to provide references and furthermore that claims are undisputed, *i.e.* always paid when requested. In this scenario, vertex v_b can purchase goods valued at up to \$150 from vertex v_s. To insure himself v_b obtains:

- a reference valued at \$100 from v_1 against bad behavior of v_s.
- a reference valued at \$50 from v_1 against bad behavior of v_2.
- a reference valued at \$50 from v_2 against bad behavior of v_s.

Similarly, to insure the transaction v_s should obtain:

- a reference valued at \$150 from v_3 against bad behavior of v_b.

Once these references are obtained, the transaction can go ahead and neither party will lose money if the other party misbehaves.

If, for example, v_s does not deliver the service/product, v_b can simply obtain the \$150 paid by contacting v_1 and v_2. If v_2 declines to pay then v_b can recover this loss by asking v_1 for a further \$50. In the case that v_1 declines to pay the \$150 then v_b's original assessment of providing a reference for v_1 in the value of \$150 was incorrect and v_b should have deposited sufficient funds upon v_1's entry to the system to cover this situation (see section 4.1).

Now, if parties are always willing to provide references and claims are undisputed then v_b can cheat by also asking for a reference from v_3 and later claiming that v_s did not deliver the product. This strategy can provide v_b with an extra \$50 at no cost. This problem will be addressed in sections 3.3 and 4.3.

In section 3.1 below, we will describe TrustDavis from the point of view of the purchaser and in section 3.2 we will consider the same problem from the point of view of the insurer.

3.1 Paying for References

Suppose that in order to provide an incentive, parties are paid for the references they provide. We view this as each party becoming an insurance broker who will sell a reference (or insurance) relating to a specific transaction. How much should each party be paid for the references they provide?

First consider the situation where the reference is provided to a party that will under no circumstance make a false claim and, thus, is *ultimately trusted*. For example, v_b could be v_1's mother. In this case, v_1 can be certain that if v_b makes a claim, then it was because v_s did not deliver the product as agreed. Although v_1's trust in v_b assures him that he will not have to pay for false claims made by v_b he has no guarantee that v_s will fulfill his end of the transaction. Thus, v_1 is still taking some risk. If v_1 takes on this risk and is not appropriately rewarded for it (as would be the case if v_b were my mother!), a sequence of bad transactions could eventually drive him bankrupt.

The criterion we use to establish how much v_1 should be paid for the references he provides is that of *no riskless profitable arbitrage*. The approach followed here was proposed in [4] for pricing stock options. The idea is as follows. We assume that v_b is only interested in the transaction because her valuation for the good being provided uS is higher than the price S at which the product is offered (*i.e.* $u > 1$). Furthermore, there are only two possible outcomes to the transaction between v_b and v_s:

- the transaction completes successfully and v_b pays S dollars and receives a good worth uS; or
- v_s delivers a product of inferior quality (or does not deliver) and v_b loses her payment but gets a product valued at dS. [4]

This situation is shown in Figure 2(a), where q and p are the probabilities that the transaction goes well or fails respectively. Figure 2(b) depicts the two possible outcomes of the

[4] Again, v_b is ultimately trusted so she *always* pays if the transaction occurs.

transaction between v_b and v_s that hold when the transaction is insured under TrustDavis. In the model, v_b pays C to obtain a reference from v_1, where v_1 agrees to pay v_b the amount K if the transaction fails. The "insurance premium" C is not recovered by v_b after the transaction is over; thus, in order to insure herself against a bad transaction, v_b must be willing to share part of the proceeds that she would obtain from a successful transaction with the insurer.

We also assume that v_b can perform riskless borrowing and lending at an interest rate of $(r-1) \times 100\%$ over the period of one transaction, under this assumption x dollars before the transaction become rx afterward. We view borrowing and lending money as selling and buying bonds (at rate r). Furthermore, buying goods from v_s can be seen as acquiring the rights to get the same goods delivered. The economic value of those rights may fluctuate and is only set once the delivery actually materializes. We can now determine an upper bound on how much v_b is willing to pay v_1 for the privilege of receiving a reference that will provide her with K dollars if v_s does not fulfill his end of the transaction.

EXAMPLE 1: Assume that v_b can borrow and lend money at a rate of $r = 1.25$. She wishes to purchase 3 shirts that are on sale at the discount price of \$50 dollars each. She has seen the very same shirts advertised for \$100 dollars at a different store and is suspicious that the items on sale are of inferior quality and in reality are only worth \$25. For a net cost of 30 dollars v_b can make sure she will not lose money in the transaction:

1. Instead of buying 3 shirts, she buys 2 and waits to buy the third later, saving \$50.

2. She adds \$30 of her own money and lends the resulting \$80, by buying a bond.

The transaction either succeeds or fails. If the transaction goes well, the shirts are worth \$100 each. She will have missed the opportunity to buy one shirt at the cheaper price of \$50. However, she will have obtained $1.25 \times 80 = 100$ dollars from her loan and she can use the money obtained to purchase the remaining shirt as desired (at \$100 each). In this case, she is able to obtain the 3 shirts for the added cost of \$30 which brings her to a total of $3 \times 50 + 30 = 180$ dollars.

If the transaction fails, the shirts are only worth \$25 dollars each. She can sell the shirts obtaining $2 \times 25 = 50$ dollars. Adding this sum to the \$100 obtained from the loan, she recovers her original $3 \times 50 = 150$ dollars she risked on the transaction[5].

We see that \$30 is an upper bound on how much v_b would be willing to pay for references to insure the transaction, since for \$30, she can insure herself as described.

We view the above example as a situation in which v_b purchases not only the shirts she wants, but also a hedging portfolio to insure the transaction. If the transaction fails the portfolio will pay K dollars, if it succeeds the portfolio's net worth will be zero. The portfolio purchased is such that the sum of both actions — buying the shirts and buying the portfolio — results in the desired outcome whether or not the transaction succeeds. If the transaction succeeds then the goods are obtained for the desired price and if the transaction fails the portfolio will pay K dollars.

In the example above, v_b was willing to pay the amount of \$50 for a good that, at the end of the transaction, would be worth either \$100 or \$25. Thus, buying goods online from v_s is very similar to buying shares in the stock market where one cannot predict the future value of those shares. With this in mind, we need to establish the composition of the hedging portfolio that will enable us to achieve the desired outcomes.

Before the transaction, the hedging portfolio is composed of Δ "shares"[6] and B bonds. Its value (cost) is $C = \Delta S + B$ dollars per item (in the example above this means per shirt). The hedging portfolio insures the transaction by providing K dollars if the transaction fails and zero dollars if it succeeds. In other words, after the transaction the portfolio must be valued at:

- $C_u = 0$ dollars, if the transaction goes well; or
- $C_d = K$ dollars, if the transaction fails.

We know that after the transaction each share S will be valued at uS if the transaction succeeds and dS if it fails. Similarly, all bonds will be valued at rB. Thus, to find the composition of the hedging portfolio we need to solve for Δ and B the equations:

$$\Delta uS + rB = C_u = 0$$

$$\Delta dS + rB = C_d = K$$

yielding,

$$\Delta = \frac{C_u - C_d}{S(u-d)} = \frac{-K}{S(u-d)} \tag{1}$$

$$B = \frac{uC_d - dC_u}{r(u-d)} = \frac{uK}{r(u-d)} \tag{2}$$

[5] Of course, she lost the "insurance premium" of \$30, but we can modify the values in the example to incorporate the premium.

[6] We view a "share" as a consummated purchase that provides the right to get an item delivered. Thus, if a party has a positive number of "shares" it has the right to receive products. On the other hand, if a party has a negative number of "shares" it has the obligation to deliver products.

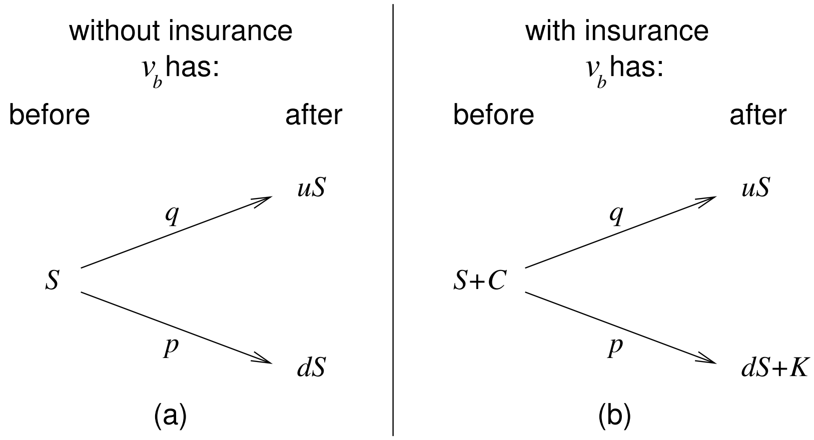

Figure 2. The two possible situations a node may face.

Thus, the hedging portfolio is composed of Δ shares and B bonds as described by the equations above and its price before the transaction is given by $C = \Delta S + B$ or more explicitly:

$$C = \frac{K}{r} \frac{(u-r)}{(u-d)} \tag{3}$$

Examining Equation 3 above provides an intuitive view of the composition of the cost of insurance in TrustDavis. The ratio K/r is simply the value of K, time corrected to the period before the transaction. The quantities u,r and d all describe *per dollar* values. Thus, $(u - d)$ is the amount risked *per dollar* in the transaction. Similarly, $(u-r)/(u-d)$ is the fraction of the total capital risked that is above the riskless interest rate. See [4, section 3] for more details. Note also that Δ is usually negative meaning that the purchaser of the portfolio should *short sell* that amount in "shares"[7].

In the example above we assumed that if the transaction falls through the buyer can recover the value $dS = \$25$ per item by selling the items after the transaction. Thus, for v_b to recover the \$50 she spent per shirt, K need only be \$25 per item. Substituting the values of Example 1 into equations 1 and 2 we obtain that the hedging portfolio for one item (*i.e.* for one shirt) has $\Delta = -1/3$ and $B = 80/3$.

3.2 Minimizing Risk

Above, we analyzed the situation from the point of view of the purchaser and obtained an upper bound on the price of a reference. Now, we look at the same situation from the point of view of the insurer and establish lower bounds on the same price.

[7]In other words, one should acquire the obligation to deliver goods at a later time. (One pays a *negative* price for acquiring an *obligation*.)

In Example 1, v_1 provides a reference. Two different circumstances may arise under which v_1 has to decide whether or not to provide a reference:

- We may have a decision problem where the price v_b is willing to pay for a reference of K dollars is already fixed (say as a percentage of the total transaction). In this case, v_1 should decide whether or not to provide such reference.

- Alternatively, v_1 may wish to place a bid to provide such a reference. In this case, v_1 needs to establish a lower bound so that it does not lose money by bidding too low and assume too much risk for the reward.

Both of these situations differ from the investment scenario we considered in 3.1 we assume v_1 has no say in *how much* of the money total available for each reference will be used. We assume v_1 faces a "take it or leave it" situation in which the buyer already knows what transactions she wishes to perform and how many items she wishes to buy, thus the transaction value is fixed. This extra constraint enables us to find precise lower bounds on the price and thus to establish whether providing such a reference is a good proposition for v_1.

To begin the analysis let us formulate the problem v_1 faces in the same way we did for v_b. This is shown in Figure 3. For each item v_1 decides to insure, he risks K dollars of his capital. In exchange, it keeps the insurance premium C. Thus, v_1 possibly obtains a return of $1 + \frac{C}{K}$ on his investment if the transaction goes well. We assume that v_1 has a fund with an initial total of W_0 dollars and is also able to estimate the probability the transaction may fail p. If v_1 risks too much money in each transaction it insures, then gambler's ruin may occur. We follow the reasoning presented in [15] to obtain an upper bound on the amount that can be

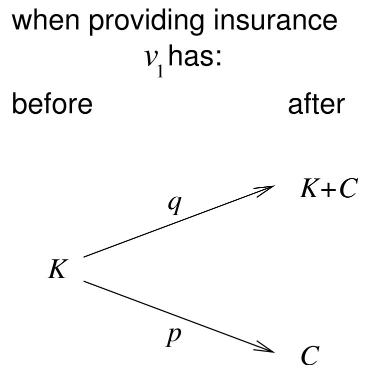

Figure 3. Insurer's point of view.

risked — or equivalently a lower bound on the price — by using the Kelly criterion [9].

The Kelly criterion assumes that each transaction can be repeated indefinitely in a sequence of rounds. Denoting by W_0 the initial capital and by W_n the capital available after round n the Kelly criterion suggests we should maximize the expected value of the growth rate of capital:

$$R = E\left\{\log\left[\frac{W_n}{W_0}\right]^{\frac{1}{n}}\right\}$$

We denote by W_{i-1} the total capital v_1 has available for insuring a particular transaction at round[8] i. Assuming v_1 risks a fraction f of W_{i-1} at round i and the transaction succeeds we have:

$$W_i = \left(1+\frac{C}{K}\right) f W_{i-1} + (1-f)\, W_{i-1} = \left(1+\frac{C}{K}f\right) W_{i-1}$$

Similarly, if the transaction fails v_1 gets to keep only the insurance premium C. In this case, the wealth after the transaction is given by:

$$W_i = \frac{C}{K} f W_{i-1} + (1-f)\, W_{i-1} = \left(1+\frac{C}{K}f - f\right) W_{i-1}$$

Thus, v_1's wealth after n rounds is given by:

$$W_n = W_0\left[\left(1+\frac{C}{K}f\right)^G \left(1+\frac{C}{K}-f\right)^B\right]$$

where G is the number of times the insured transaction succeeds (good) and B is the number of times the insured transaction fails (bad). Obviously, $G + B = n$. Calculating the expected value of the growth rate coefficient we have:

$$R(f) = E\left\{\log\left[\frac{W_n}{W_0}\right]^{\frac{1}{n}}\right\}$$

$$= E\left\{\log\left[\left(1+\frac{C}{K}f\right)^G \left(1+\frac{C}{K}f - f\right)^B\right]^{\frac{1}{n}}\right\}$$

$$= E\left\{\frac{G}{n}\log\left(1+\frac{C}{K}f\right) + \frac{B}{n}\log\left(1+\frac{C}{K}f - f\right)\right\}$$

$$= E\left\{\frac{G}{n}\right\}\log\left(1+\frac{C}{K}f\right) + E\left\{\frac{B}{n}\right\}\log\left(1+\frac{C}{K}f - f\right)$$

$$= q\log\left(1+\frac{C}{K}f\right) + p\log\left(1+\frac{C}{K}f - f\right)$$

Solving the above equation numerically for $R = 0$ and different values of p yields the minimum values for the ratio $\frac{C}{K}$ shown in the graph of Figure 4.

Note that if v_1 receives a value C that yields smaller ratios than the ones shown then the growth rate is negative and gambler's ruin will occur. Alternatively, if the price C provides larger returns the insurer's capital will grow at a rate[9] R.

3.3 Dealing with False Claims

Up to this point we have considered situations where the party receiving the insurance is considered *ultimately trusted*: There were no false claims. We will call this the *no false claims* scenario, NFC. Now we take into account the possibility that the insured party may cheat and stake an undue claim that the insurer has to pay. Similarly, we call this the *false claims* scenario, FC.

The analysis in 3.2 was done in the NFC scenario. We considered a successful transaction one in which the insurer kept the money K and the premium C. A failed transaction is one in which the insurer has to pay K dollars. Because we made no assumptions about the reasons a transaction may fail, the same analysis still holds under FC. We only need to change the probability that the transaction may fail p to reflect the new risks.

4 Strategies

In this section, we present a strategy for trading online and providing references that enables an honest individual to limit how much damage a malicious collusion of players can do. In all cases, we assume that a potential trader will

[8] If v_1 has a separate sub-fund of total capital W_{i-1} for each party v_s and is using the strategies describe in section 4 then v_1 cannot be successfully exploited by a malicious party but the overall growth coefficient for the sum of all sub-funds may be smaller than it would be if the same fund is used for all parties and defaults are random events. See chapter 15 in [3].

[9] The minimum growth rate desired can be set to a value large than zero such as the "zero risk" interest rate.

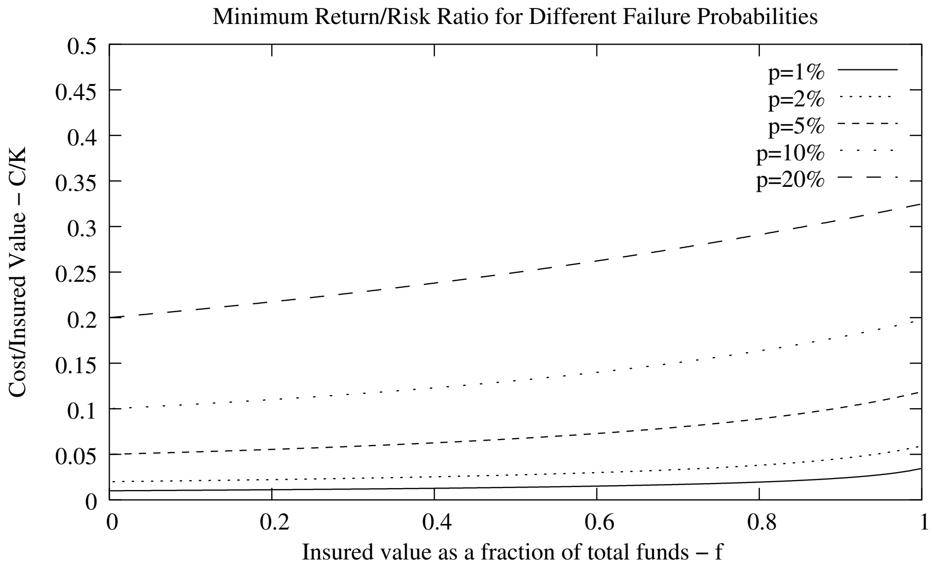

Figure 4. Minimum cost of a reference as a function of the funds available and the probability of failure.

only engage in trade if his valuation for the goods being bought (or sold) is larger than the opportunity cost of the transaction.

EXAMPLE 2: Assume that v_b has \$190 to spend and is considering buying a few gifts online. She narrows down her search to 3 good deals. She can:

1. Buy 3 shirts for \$50 each, from an *unreliable* source v_s insuring the transaction for \$40. She thinks each shirt is worth \$100.

2. Buy 2 pairs of shoes for \$70 each, from a reliable retailer. She thinks each pair is worth \$90.

3. Buy 1 game console for \$150, also from a reliable online shop. She thinks the console is worth \$240.

Assuming that money leftover is not spent, if v_b chooses alternative 1 and the transaction goes well, she will have obtained 3×100=300 worth of goods for 3×50+40 (insurance)=190 dollars. Choosing option 2 she will have 2×90+50 (leftover cash)=230 dollars worth (in goods and cash) for the same 2×70+50=190 dollars. Similarly, if she chooses alternative 3 she will have 240+40 (leftover cash)=280 dollars worth. Clearly, her best option is to buy the shirts. We consider the opportunity cost of that transaction to be the value of the second best option, \$280.

In the example above, if the transaction goes well, v_b obtains an extra 300-280=20 dollars through trading with v_s that she would not have obtained had v_s not been available. Furthermore, because the transaction was insured, v_b did not risk any money to obtain the extra \$20[10]. Under these conditions we suggest that to insure herself for future transactions v_b should save \$5 of the \$20 obtained in a fund $F_b(v_s)$ that will provide references to v_s. In doing so, v_b is extending v_s a credit line — a fairly common business practice.

4.1 A Strategy When There Are No False Claims

We can now describe a non-exploitable strategy for trading and providing references online. We first consider the NFC scenario. To avoid exploitation v_b proceeds as follows:

1. During the initialization step, v_b only provides references to agents she trusts and that will not default on their obligations. She can also provide references to agents that leave a security deposit under her control.

2. v_b only engages in insured transactions by obtaining references for them through the individuals she trusts.

3. After every transaction (buy or sell) v_b saves part of the gains obtained *in excess of the opportunity cost* in separate funds that are linked to each trade partner (to

[10] We assume the insurance was such that she would also receive \$280 dollars if the transaction failed. A similar argument can be made if v_b receives \$190 in insurance in case the transaction fails, but can still buy the console after receiving the insurance money.

provide references for them). This helps her to insure future transactions with each partner.

4. v_b provides references to others by charging premiums as described in section 3.2. This provides some confidence that the money saved will grow at an specified rate. Again, each premium received is put in a separate fund that is linked to the agent it insured against bad behavior of (*not* the agent who paid for it).

Because v_b only engages in insured transactions, this strategy limits v_b's exposure to the total amount in the funds $F_b(v_i)$ for all v_i she is willing to provide a reference for. This value, $T_b = \sum_i F_b(v_i)$, is only changed by adding money earned through trading *in excess of the opportunity cost* or through selling references. In either case, the funds have been obtained through the trading in TrustDavis and from the parties they may benefit.

4.2 An Alternative Algorithm for Obtaining References

In the NFC scenario, all references are directly obtained by the party being insured, v_b. Thus, intermediate nodes will pay insurance, in the event of a failed transaction, directly to the insured party. In this scenario, the insured party v_b is the same for all intermediate nodes. Intermediate nodes also provide insurance against bad behavior of other intermediate nodes. Thus, the party being insured, v_b, insures a transaction by walking the paths from itself to the party they wish to trade with, v_s, as described in section 3. We assume some efficient distributed algorithm is used to find such paths. This procedure makes price negotiation easy as all communication occurs between the agent asking for the references and the agents providing references with no intermediaries.

In the FC scenario, the above algorithm cannot be used because some insurers may no longer trust the party asking for the reference. We change the algorithm, described in section 3, as follows: when v_b wants to make a purchase from v_s she only asks her neighboring nodes to provide references for the transaction. Her neighbors, in turn, ask for references on their own behalf from their neighbors along a path from themselves to v_s. Once those references are established and v_s is reached, the replies propagate back to v_b. In Figure 1 this corresponds to the following sequence of requests and replies:

1. v_b asks v_1 for a reference valued at \$150 against bad behavior of v_s and waits for a reply.

2. v_1 asks v_2 for a reference valued at \$50 against bad behavior of v_s and waits for a reply.

3. v_2 verifies that he "trusts" both v_1 and v_s more than \$50 and replies to v_1 providing the reference.

4. v_1 verifies that he has at least \$150 dollars of "flow capacity" to both v_b and v_s and replies to v_b providing a reference.

5. v_b goes ahead with the transaction.

Note that by using this algorithm the beneficiaries of the references provided are always neighboring nodes and more complicated price negotiation is required, because the party paying for the insurance is not in direct communication with the insurers.

4.3 A Non-Exploitable Strategy

Let us call a party against whose misbehavior a reference is provided the *object* party. Also, let us call the party that receives money if a transaction fails the *insured* party. The party providing the reference is the *insurer*.

Consider applying the strategy described in section 4.1 above to the example in Figure 1, using the algorithm described in section 3 in the NFC scenario. When v_1 provides a reference to v_b against bad behavior of v_s, v_1 limits his liability to the "capacity" of the *edge* from v_1 to v_s, *i.e.* \$100. Similarly, v_1 is also asked to provide a reference against bad behavior of v_2 and he also limits his liability to the "capacity" of the *edge* from v_1 to v_2, \$50. So v_1's total liability for the transaction is \$150.

In the FC scenario, the outcome of the transaction no longer relies only on the trustworthiness of v_s, it also depends on v_b. If v_1 provides references to v_b with a total value that is smaller than the total "network flow capacity" from v_1 to v_b[11], then v_1 can recover potential losses caused by v_b by withdrawing money from the appropriate funds. In the example, this is $F_1(v_b)$. Before doing so, v_1 must perform due diligence and establish that the transaction failed due to v_b and not v_s. If the transaction fails due to v_s, then v_1 can only recover \$100 from his own funds. Thus, if v_1 is to provide the same total liability of \$150 he should obtain from v_2 a reference *for himself* against bad behavior of v_s. By obtaining this reference from v_2, v_1 simply limits his liability to the smallest of the two "flows" v_1 to v_b and v_1 to v_s. This is the strategy we propose should be used. It requires the algorithm described in section 4.2.

In the FC scenario, when an insurer provides a reference, that reference will be unclaimed only if both the insured and the object party behave appropriately. Therefore, it is too restrictive to deposit the premium received for providing this reference in a fund linked exclusively to the name of the object party as in the NFC scenario. This is because

[11] Note that the total network flow capacity from v_1 to v_b in Figure 1 is \$300 but each path can only be used once in the following discussion. Thus, we only consider the edge (v_1, v_b) as a return path to v_b, since (v_1, v_s) and (v_1, v_2) are already being used in the forward direction to link to v_s.

the insurer cannot be exploited by linking this fund to either party and linking it to only one party unnecessarily limits the number of transactions the insurer can be involved in. A more flexible approach is to have yet another fund that can be used to provide references to *either* party.

In summary, to adapt the strategy proposed in section 4.1 to the FC scenario participants need to perform three actions differently:

- As described in section 4.2, the insurer must acquire insurance (by obtaining references from his neighbors) when asked to insure a transaction "flow" that is greater than his direct capacity to insure, as measured by the capacity of the edge from him to the object party.
- When providing references, the insurer has to limit his liability to the minimum of the two "flows"
    1. from the insurer to the insured party;
    2. from the insurer to the object party.
- Link money received through selling references not only to the object party but to the pair (insured party, object party).

5 Conclusion

In this paper we proposed a reputation system with the following four important properties:

- Honest participants can limit the damage caused by *malicious* collusions of dishonest participants.
- Malicious participants gain no significant advantage by changing or issuing themselves multiple identities.
- There is strong incentive for participants to provide *accurate* ratings of each other.
- It requires no centralized services, and thus can be easily distributed.

We think interesting directions for future research are to explicitly address issues that may arise within the framework due to the time varying value of money and to analyze protocols for price negotiation and finding paths distributively.

6 Acknowledgments

Dimitri would like to especially thank Matt Franklin for many insightful conversations.

References

[1] Z. Abrams. R. McGrew and S. Plotkin, "Keeping Peers Honest in EigenTrust", in *Proceedings of the Second Workshop on the Economics of Peer-to-Peer Systems*, 2004.

[2] S. Buchegger and J. Y. Le Boudec, "A Robust Reputation System for P2P and Mobile Ad-hoc Networks", in *Proceedings of the Second Workshop on the Economics of Peer-to-Peer Systems*, 2004.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.

[4] J. C. Cox, S. A. Ross and M. Rubinstein, "Option Pricing: A simplified Approach", *Journal of Financial Economics*, 7 (1979), 229-263.

[5] C. Dellarocas, "Building Trust On-Line: The Design of Robust Reputation Mechanisms for Online Trading Communities", chapter VII in *Information Society or Information Economy? A combined perspective on the digital era*, edited by G. Doukidis, N. Mylonopoulos and N. Pouloudi, Idea Book, 2004.

[6] E. J. Friedman and P. Resnick, "The Social Cost of Cheap Pseudonyms", *Journal of Economics & Management Strategy*, vol. 10, issue 2, 2001, 173-199.

[7] J. Golbeck, B. Parsia and J. Hendler, "Trust networks on the semantic web", in *Proceedings of Cooperative Intelligent Agents* 2003, Helsinki, Finland, August 2003.

[8] S. D. Kamvar, M. T. Schlosser and H. Garcia-Molina, "The EigenTrust Algorithm for Reputation Management in P2P Networks", in *Proceedings of the Twelfth International World Wide Web Conference,* WWW2003, Budapest, Hungary, 20–24 May 2003. *ACM, 2003.*

[9] J. L. Kelly, "A new interpretation of information rate." *Bell System Technical Journal*, 25 (1956), 917-926.

[10] P. Kollock, "The Production of Trust in Online Markets." in *Advances in Group Processes,* Vol. 16, edited by E. J. Lawler, M. Macy, S. Thyne and H. A. Walker, Greenwich, CT, JAI Press, 1999.

[11] S. Lee, R. Sherwood and B. Bhattacharjee, "Cooperative Peer Groups in NICE", IEEE Infocom, April 2003.

[12] P. Resnick, R. Zeckhauser, "Trust among strangers in internet transaction: Empirical analysis of eBay's reputation system." In *Working paper for the NBER Workshop on Empirical Studies of Electronic Commerce,* 2000.

[13] P. Resnick, R. Zeckhauser, E. Friedman and K. Kuwabara, "Reputation Systems", *Communications of the ACM*, Vol. 43, No. 12 (Dec., 2000), 45–48.

[14] T. Riggs and R. Wilensky, "An Algorithm for Automated Rating of Reviewers", in *Proceedings of the First ACM/IEEE-CS joint conference on Digital libraries*, 2001.

[15] L. M Rotando an E. O. Thorp, "The Kelly Criterion and the Stock Market", *The American Mathematical Monthly*, Vol. 99, No. 10 (Dec., 1992), 922–931.

[16] B. Yu and M. P. Singh, "Distributed Reputation Management for Electronic Commerce", *Computational Intelligence*. Vol. 18, No. 4, November 2002, 535-549.