Modeling Trust Using Transactional, Numerical Units

Reid Kerr and Robin Cohen
David R. Cheriton School of Computer Science
University of Waterloo
{rckerr, rcohen}@cs.uwaterloo.ca

Abstract

In large electronic marketplaces populated by buying and selling agents, repeated transactions between traders may be rare. This makes it difficult for buying agents to judge the reliability of selling agents, discouraging participation in the market. A variety of systems have been proposed to help traders to find trustworthy partners; however, most proposed systems suffer from multiple vulnerabilities that might be exploited by unscrupulous parties. In this paper, we propose a new model, wherein abstract units are used to represent trust in much the same way that units of money represent value. In a manner similar to money, 'trunits' flow during transactions. A trader's trunit balance determines if they are trustworthy for a given transaction. Faithful execution of a transaction results in a larger trunit balance, permitting the trader to engage in more transactions in the future—a built-in economic incentive for honesty. We demonstrate that for a wide range of realistic market parameters, the Trunits mechanism ensures that honest sellers profit more than dishonest sellers. We also discuss how intrinsic properties of our model make it secure from many of the attacks to which other systems are vulnerable. In summary, we present our Trunits model as the basis for modeling trust in electronic marketplaces, useful as buying agents develop algorithms to intelligently choose trustworthy sellers for their business partners.

Keywords

Trust and reputation, multiagent systems, e-commerce marketplaces, buying and selling agents

1. Introduction

While trust and reputation have received a great deal of attention recently in a variety of areas, e-commerce scenarios have seen particular emphasis.

New technologies have allowed the development of new types of marketplaces, with very different characteristics from those of the past. These marketplaces may involve extremely large numbers of buyers and sellers; very few of these traders will know one another, and transactions are often of a one-time variety, where the buyer and seller may never do business together again. eBay is the best known example of such a marketplace, and is discussed often in the research.

Since traders in such markets do not have much opportunity to form relationships with one another, trust does not develop as it might in a more traditional commerce setting. Further, since each trader has contact with such a small fraction of the other traders, development of reputation via traditional 'word-of-mouth' is inhibited—it is unlikely that someone you know has knowledge of the trader you wish to evaluate.

Despite these obstacles, it is critical for the success of such marketplaces that traders be able to 'trust' one another—if they cannot, they simply won't trade in the markets. Thus, much research attention has been devoted to developing methods and models that are effective in this type of scenario.

Many of the proposals in this area consist of complex models, which require the gathering, maintenance, and processing of a great deal of information. When faced with such complexity, it is natural to wonder if perhaps a more 'natural' model might be found that simplifies the issue.

In pondering the problem, we were inspired by the concept of money to consider trust as an abstract numerical quantity, one that flows during the course of transactions. In order to develop our model, we first examined several existing approaches to the modeling of trust and reputation in electronic marketplaces, and developed a catalogue of vulnerabilities that would be beneficial to address within our own approach. In the sections that follow, we briefly summarize relevant related work, and then present our own Trunits model in detail.

2. Survey of Existing Approaches and Issues

An enormous amount of work has been conducted in this field, and it is well beyond the scope of this paper to examine all of it. Instead, we restrict our survey to a small sampling of models that share at least one key characteristic with our proposed approach:

- They are identified as applicable to marketplaces;
- They rely on global measures of trust or reputation;
- They measure trustworthiness using a 'transactional' or 'accounting' system, which tracks (unbounded) quantities rather than ratings.

2.1. The eBay Reputation System

eBay [5] typifies the type of 'new market' with which many investigators are concerned, and has demonstrated considerable success in gaining the confidence of traders. A discussion of eBay's system is a valuable starting point, for two main reasons:

- It provides a convenient framework within which to discuss the issues faced by trust/reputation systems for marketplaces, with real world context, and
- It provides a benchmark against which other systems might be measured. Given that eBay has an established, implemented system, any proposal that specifically targets markets of this type must improve upon eBay's system in some way to be interesting. (Here, we mean 'improve upon' in a broad sense; a system might be interesting for a variety of reasons, without necessarily 'outperforming' that of eBay.)

The eBay system is fairly simple in operation. Before engaging in a transaction, both buyer and seller have the opportunity to view each other's 'feedback profiles' (discussed below). After a transaction has been completed, each trader has the opportunity to give feedback on the other, which is used to update the other's profile. Feedback consists of a rating of 'positive', 'neutral', or 'negative' (and optionally, a free-form text comment).

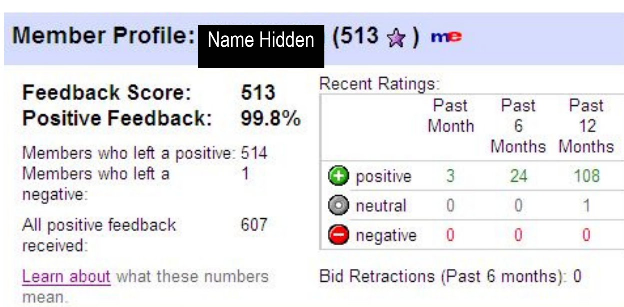

Figure 1: A feedback profile

A feedback profile is shown in Figure 1. The profile presents two key pieces of information about the user:

1. The *feedback score*, the number of positive ratings less the number of negative ratings. This score seems intended to give a sense of the volume of trustworthy transactions in which the user has engaged.
2. The *positive feedback* percentage, the number of positive ratings divided by the total number of ratings. This score seems intended to give a sense of the likelihood that the user will be trustworthy in a given transaction.

(In both cases, only ratings from unique users are considered.) Based on this information, the user may decide if she considers the potential partner trustworthy enough to trade with him.

2.2. Possible problems with the eBay System

As the level of participation shows, eBay has been successful in convincing potential traders that the system offers adequate safety to join in. The question of satisfaction within the marketplace is a different issue, however. Superficially, the mechanism is extremely successful, with 99.1% of feedback being positive [4]. A deeper examination, however, suggests that the data does not accurately reflect real satisfaction levels. In [1], it is noted that in systems such as eBay's, a trader who receives a bad rating may retaliate with (undeserved) bad feedback. Fear of such retaliation discourages negative feedback, artificially inflating rankings. A study cited in [13] found that there is a high correlation between the ratings of buyers and sellers in transactions, implying that feedback may say more about the smoothness of the bilateral execution of the transaction, than about the reliability of a single participant. In [2], the author cites Resnick and Zeckhauser's findings that there is often pressure on traders to give positive ratings.

Aside from the validity of ratings, there are also questions about whether users can effectively and accurately make use of the reputation data. One problem, discussed in [4], is that users are required to interpret feedback profiles themselves, without any context or information about the users providing the feedback. Reputation impacts both the likelihood and price of sale, but not consistently so, indicating that individual buyers interpret the reputation data very differently. Thus, it is hard to measure the value of reputation, and hard for users to make decisions based on such value.

2.3. Vulnerabilities in the eBay System

Beyond these more abstract problems, there are specific vulnerabilities within the eBay reputation system that dishonest sellers can use to take advantage of buyers. According to [1], studies have shown a substantial amount of fraud in electronic marketplaces, substantiating the existence of weaknesses.

Several key vulnerabilities are identified below. (The names assigned to the problems are our own, and are not standard terminology.)

The 'Reputation Lag problem'

In many marketplaces, buyers are required to pay for goods before the seller delivers them. After payment, there is usually some delay before the buyer actually receives the good (due to processing, shipping, etc.), and has the opportunity to evaluate the transaction and/or give feedback. For example, on eBay the lag between payment by the buyer and registering the buyer's feedback might range from a few days to a few weeks.

This lag opens a window of opportunity for a seller to engage in virtually unlimited cheating. Consider a seller who decides to cheat a buyer on a transaction; the situation is depicted in Figure 2. The seller knows that he intends to cheat from the time of sale, but the buyer will not know until some time later (after he receives an inferior good, or gives up waiting for a good that never arrives). Due to the lag in the buyer's negative feedback being posted, other buyers will not be alerted to the cheater's dishonest behaviour until some time after it begins. Thus, for the entire duration of the lag, the seller can make use of his good reputation to cheat a virtually unlimited number of buyers. While the seller's reputation will eventually suffer (and likely hinder future sales), the profits earned from unbounded cheating during the window can exceed those lost in the future.

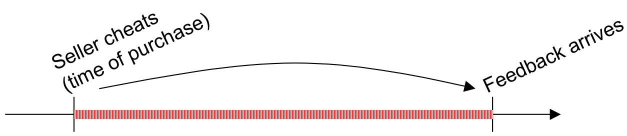

Figure 2: The window of opportunity resulting from the reputation lag problem

The 'Value Imbalance problem'

In [3], the author identifies the 'value imbalance' vulnerability. This vulnerability stems from the fact that in many systems (including eBay), the impact of each piece of feedback is not related to the value of the transaction—feedback on very small transactions is weighted equally with feedback from very large transactions. A dishonest seller can take advantage of this property to build up reputation by honestly executing a number of small-value trades, and then using the accumulated reputation to cheat a seller in a very high-value transaction.

For example, if the seller were to engage in 5 honest transactions, each of $1 in value, his feedback score would be 5, with a feedback percentage of 100%. To potential buyers, he appears to be quite honest. Using this honest appearance, he might lure a buyer into making a $1,000 purchase, and cheat by not delivering the product. Worse still, after receiving the negative feedback, the seller would still have a feedback score of 4, and feedback percentage of 83.3%—his reputation is far from being irreparably damaged.

The 'Ballot-Stuffing problem'

Trust/reputation has value—good reputation is likely to result in greater sales volume, and may allow higher prices to be charged. Further, good reputation can be used to lure buyers into transactions in which they can be cheated.

Many authors (e.g. [1], [4], [14]) identify the opportunity for buyers to engage in 'ballot-stuffing'. Ballot-stuffing is collusive behaviour in which parties engage in fake transactions in order to artificially inflate their reputations.

Some systems (such as eBay) have attempted to limit the ability of users to ballot-stuff by only counting feedback from unique users. However, the vulnerability remains, due to the ease with which new identities can be created; it is trivial to create many new accounts, allowing ballot-stuffing transactions to be generated by 'unique' users.

The 'Re-entry problem'

The ease of creation of new accounts introduces another vulnerability, identified by numerous authors (e.g. [1], [4], [14]). In many systems (such as eBay), it is preferable to have no reputation than to have a bad reputation; while users may be hesitant to deal with a trader with no history, they are more hesitant to deal with a disreputable user. A seller can shed his disreputability by simply creating a new account, freeing himself of the restrictive reputation. In essence, the seller has improved his reputation without engaging in even a single honest transaction. The seller can repeat this cycle as many times as desired, cheating until his reputation is destroyed, and then beginning again with a new identity.

The 'Exit problem'

The 'exit problem' occurs when a seller who has gained some positive reputation decides to leave the market. The seller has no further need of her good reputation, so the entire accumulated reputation can be used to cheat buyers. Eventually, her reputation deteriorates to the point that no buyer will deal with her, so she completes her exit.

This is an extremely difficult problem to address, and one that afflicts most trust/reputation systems. If a seller has no further use for her good reputation, how can you prevent her from expending it by cheating others?

Unfortunately, these vulnerabilities in the eBay system also exist in many of the trust/reputation systems that have been proposed. Ideally, a system would prevent each of these problems, without introducing substantial new ones. Thus, we will use these problems as a basis for evaluating other systems, including Trunits.

2.4. A survey of selected trust/reputation systems

In the survey of trust/reputation systems conducted in [10], the authors characterize systems in a number of dimensions. We use one of these dimensions, that of 'information sources', to group systems for convenience of analysis.

Direct Experience models

Direct experience models are those in which an agent evaluates the trustworthiness of a potential partner based on the agent's own experience with that potential partner. (Here we consider only models that exclusively use direct information.)

So-called 'general trust' models like those described by Marsh [8] and Griffiths [6] can be applied to marketplaces—Griffiths specifically discusses this scenario in his paper. Both Marsh's 'situational trust' and Griffith's 'multi-dimensional trust' build on Marsh's general trust model. In this model, an agent's trust of another is reflected by a score in the range [0, 1]. With each transaction, the agent updates its trust of the other, increasing it in the case of a trustworthy transaction, and decreasing it otherwise; when faced with a choice between potential partners, the one with the highest score is selected.

In the work of Tran and Cohen ([11], [12]), both buyers and sellers are learning agents—buyers learn to maximize the expected value of the goods they purchase, while sellers learn to optimize values such as quality and price in order to maximize profit. Buyers maintain sets of sellers whom they consider to be reputable, disreputable, or of unknown reputability. In addition, buyers derive an expected value function for each seller, reflecting the estimated value of receiving good *g* at price *p* from that seller. In order to purchase a good, a buyer selects a (non-disreputable) seller providing the highest expected value; after the transaction is completed, the expected value function and sets of sellers are updated based on the experience. It is important to note that occasionally, the buyer seeks to explore the marketplace, randomly selecting a seller who is not disreputable in order to broaden the range of suppliers. The model of Tran and Cohen also addresses the important problem of value imbalance, since the value of a purchase does in fact influence how the reputability of the seller is adjusted.

As a group, these models have a common disadvantage: a buyer must learn which sellers to trust and which to distrust. This means that buyers are vulnerable to unscrupulous sellers until they have learned to avoid them, providing the dishonest sellers with an 'initial window of opportunity' to cheat. The models attempt to minimize this problem by favouring known reputable sellers, but this is only possible when a known seller offers the desired good—this occurs infrequently in the type of market with which we are concerned.

With reference to the problems identified in the Vulnerabilities section, the models of Marsh and Griffiths are subject to the value imbalance problem. All of the models tend to be resistant to the reputation lag and the ballot-stuffing problems (since they only rely on direct experience). However, all of them are subject to the re-entry problem (since unknown sellers are treated differently from disreputable ones) and the exit problem. Further, the re-entry problem can combine with the 'initial window of opportunity' in a dangerous way: a seller may re-enter the market each time that buyers learn of his dishonesty, in order to take advantage of the initial window repeatedly.

Witness Information models

Witness information models are those in which the agent makes use of information supplied by other agents in evaluating the trustworthiness of a potential partner (often supplemented by personal experience).

In [14] two systems are introduced, Sporas and Histos. Under the Sporas system, each agent has a single, global reputation score, which is updated after each transaction. According to [10], Sporas is an 'evolved version' of an eBay-like system, making use of a somewhat-related mechanism. However, it incorporates several interesting features:

- To minimize the re-entry problem, new users do not begin with better reputations than disreputable users.
- To minimize the impact of ballot stuffing, when an agent rates another more than once, only the most recent rating is considered.
- To combat the creation of new accounts for the purpose of ballot stuffing, more weight is given to ratings from users with established reputations.

The Histos system allows for personalized ratings by making use of a 'web of trust' model to obtain and evaluate recommendations, where transitivity of trust relationships is used to evaluate recommendations provided by agents with whom there is no direct relationship. Since Histos relies on paths between traders, Sporas is used in the absence of such a path.

As indicated above, Sporas is resistant to ballot-stuffing and provides no incentive for re-entry. However, both models are vulnerable to the reputation lag problem (since they rely on information provided by others), the value imbalance problem (since reputation changes are not linked to the value of transactions), and the exit problem.

Like the model of Griffiths outlined in [6], the REGRET system [9] takes a multi-dimensional perspective on the modeling of reputation. In this model, each agent maintains a database of 'impressions' of previous transactions. Agent A's individual, subjective view of B's reputation is the average of its ratings of B in applicable situations (i.e., from impressions that match the desired transaction parameters) weighted to place greater emphasis on more recent transactions. The number and variability of the ratings are used as measures of reliability of the estimate. The REGRET system also allows an agent to incorporate the views of other agents, based on group membership. However, in discussing the marketplace scenario, the authors indicate that the 'group' opinion is ultimately a global reputation value. Under these circumstances, REGRET does not diverge from the eBay system in any way that substantially addresses the problems identified in the Vulnerabilities section. (REGRET's use of direct experience as well as witness information provides no additional protection, since the use of direct experience is implicit in the eBay system—traders are aware of their own experience with other traders, and incorporate this into decisions along with the reputation data.)

The model presented by Yu and Singh in [13] makes use of referral networks, in which each agent may seek recommendations from others. In answering a request, an agent may provide one from its own experience, or may forward the query to a trusted neighbour. While many systems combine recommendations from agents, the authors have chosen to do so using the Dempster-Shafer theory of evidence, believing it to be superior in its ability to distinguish between disreputability and uncertain reputability. This model has some vulnerability to ballot-stuffing and reputation lag (to the degree that an agent relies on the information of others). It appears to be vulnerable to the re-entry problem (since it explicitly treats those with no reputation and those with poor reputation differently), the value imbalance problem (since reputation changes are not linked to the value of transactions), and the exit problem.

Transactional models

The Trunits model is difficult to classify into the information source categories identified above. As detailed below, Trunits makes use of units of trust, which 'flow' during the course of transactions. Under this model, a buyer does not form a belief of a seller's trustworthiness based on personal experience, nor does he consult with others to gain their opinions. Instead, the system provides a strong incentive for buyers to behave honestly—buyers, in turn, trust the system rather than individual sellers.

Since the trunits model treats trust as a transactional quantity, it is pertinent to consider models that do likewise. To date, we have examined few such models, but touch on work in one related area, that of peer-to-peer systems.

Peer-to-peer file sharing systems work well only if sufficient numbers of users provide content. These systems can be plagued with 'free-riders' who download large quantities of files without serving any content. Further, attacks have been made on such systems by users serving useless content disguised as more desirable material. For these reasons, there has been significant interest in trust/reputation in the peer-to-peer community, with some proposals being transactional in nature.

Under one such proposal [7], reputation scores are maintained using one of two mechanisms. Under the debit-credit policy, reputation is increased when a user contributes resources, and decreased when he consumes resources. (A credit-only version ignores consumption; the authors note its vulnerability to ballot-stuffing.) While this proposal is described as a reputation system, a user's ability to download is directly tied to the value of its contributions, meaning that uploads are effectively being exchanged for downloads—the 'reputation score' does not reflect trustworthiness, but rather takes the role of currency, representing stored value earned through transactions with others.

3. Trunits

The 'Trunits' model was inspired by the concept of money. Before the advent of money, goods and services were exchanged by bartering. This placed several limitations on trade; here, the most relevant was the requirement for buyers and sellers to interact directly, to exchange goods of comparable value. A primary function of money is to overcome this requirement.

Money is an abstract 'substance', representing quantities of value. Money flows in a transaction, mirroring the flow of value in a barter transaction: the value of the money stands in for the value of a good. Money frees the traders from the requirement that goods move in both directions—value gained from one trader can be 'spent' with another.

The key problem with trust in our new breed of marketplaces is that buyers and sellers usually do not have direct relationships, so trust cannot form naturally. Since we seek to overcome the requirement for a direct relationship—to allow trust gained from one trader to be 'spent' with another—it seems natural to consider the use of abstract trust units, or 'trunits', to play the same basic role in which money has been so successful.

3.1. The Trunits Model

As with money, the movement of trunits should mirror that of trust in a direct relationship. This movement, however, is very different from that of value. While the flow of value is an exchange process, we see the 'movement' of trust as a *risk* process, and suggest a model based on this view. We focus on trust of the seller as the primary issue:

- Before a buyer will purchase something from a seller, the buyer must have sufficient trust in the seller. The degree of trust required is dependent on a number of factors; the price of the item is likely a major one.
- After purchasing the good, the buyer will evaluate it.
    - If the good met her expectations (i.e., it was at least as good as was advertised by the seller), then the seller is likely to gain more of her trust.
    - If the good did not meet her expectations, then the seller is likely to lose some of her trust.

Based on this view, we suggest a model that makes use of abstract units of trust, and in which trust of the seller is not tied to a specific buyer:

- The seller has some quantity of trunits, representing all of the trust gained from all buyers to date. For a buyer to consider purchasing something from a seller, the seller must possess a sufficient degree of trust,

i.e., must hold sufficient trunits. The required number of trunits is tied to the price of the good.
- After purchasing the good, the buyer will evaluate it, relative to her expectations.
  - If the good met her expectations, then the seller gains some additional quantity of trunits.
  - If the good did not meet her expectations, then the seller loses some quantity of trunits.

The number of trunits gained is proportional to the size of the sale. Honest execution of small transactions will allow a seller to continue making small sales, and to grow his sales volume, but will not allow him to immediately jump to disproportionately large sales for which he has not demonstrated trustworthiness.

In the sections that follow, we consider a very simple mechanism built upon this model.

3.2. The Basic Trunits Mechanism

When an agent wishes to make a sale, we require him to put up a quantity of trunits to 'cover' the sale. (We discuss the issue of how a seller initially obtains trunits later in this paper.) These trunits represent the trust that the seller is risking by engaging in a transaction. We require that the number of trunits risked be directly tied to the value of the transaction, using the formula:

$$V = rT$$

where V is the value (selling price) of the transaction, T is the number of trunits, and r is the required *risk ratio*. The trunits are put into escrow with the market operator, pending completion of the transaction.

Upon completion, if the buyer rates the transaction as unsatisfactory, then the seller loses the T trunits placed in escrow. If, on the other hand, the buyer rates the transaction as satisfactory, then the T trunits are returned to the seller, along with some additional quantity of trunits related to the value of the transaction, for a total of:

$$(1+p)\ T = (1+p)\ V/r$$

where p is a *premium* or *reward* of additional trust for acting in an honest manner. (Since the amount of trust in a marketplace is not fixed, the market operator can create or destroy trunits as required, and furnishes the reward trunits in the case of an honest transaction.) It is suggested that p be less than 1, in order for trust to be harder to earn than it is to lose (as suggested in [11], [12]), but this is not a strict requirement. (In the basic mechanism presented here, the same values of r and p are used for all traders and transactions.)

From a buyer's perspective, no evaluation or computation is required prior to purchasing to determine if a seller is trustworthy—if the seller possesses enough trunits for a transaction, then *by definition*, she is trustworthy *for that transaction*. The market operator will not allow a transaction to be executed unless the seller has sufficient trunits.

An Example

To illustrate the mechanics of the system, consider the following example:
- A seller has 20 trunits.
- The required ratio $r = 5$.
- Since $V = rT$, the maximum transaction value that the seller can cover is 20 * 5 = $100.

Suppose now that the seller engages in a sale for a price of $50.
- Since $V = rT$, the seller must place 50/5=10 trunits in escrow.
- If the buyer is unsatisfied, the seller loses the 10 trunits in escrow. She now has 10 trunits remaining, meaning that she can now cover transactions with a maximum value of $50.
- Suppose the reward ratio $p = 0.2$. If the buyer is satisfied, then the seller's trunits are returned to her, and she receives an additional $pT = 0.2(10) = 2$ trunits. She now holds a total of 22 trunits, and can cover transactions with a maximum value of $110.

Important properties of the model

The intention, illustrated in this example, is that trunits themselves will be valued, deriving from the fact that they enable profitable transactions. Under this mechanism, the more trustworthy a seller becomes the greater volume/value of sales she can execute. One possible downside of this policy is that it restricts the total volume of trade, since sellers may not possess enough trunits to sell their desired volume of goods. However, it has the beneficial consequence that trustworthy traders tend to dominate the market due to the larger sales volume they can achieve. In many scenarios, this is a favourable trade-off.[1]

An obvious question is, if trunits have value, how does this system differ from one in which the seller puts up a cash bond for each transaction, which is lost in the event of dishonesty? There are several key issues that highlight the differences between the two mechanisms:
- If the seller puts up a cash bond, and is determined to have been untrustworthy in executing the transaction, what happens to the bond? If it is paid out to the buyer, then the buyer has a financial incentive to rate the seller as untrustworthy even when the seller has been honest. If the market operator keeps it, then the market operator has an incentive to structure the system to encourage unfairly negative ratings, to the detriment of sellers.
- How large does the bond have to be for a transaction? If the bond is not at least as large as the seller's

[1] Discussion of 'market dominance' can raise concerns of monopolistic behaviour. However, it should be noted that, at least in the case of the fixed risk-ratio, a seller cannot use a large trunit balance to make its offering more attractive than that of other sellers; any seller with sufficient trunits to cover a transaction can compete. Price-based monopolistic behaviour is an orthogonal issue to that of trust, and falls outside the scope of our model.

cost to furnish the purchased good, then the seller realizes greater profit from simply cheating the buyer (by keeping the money without providing the good) than from honestly completing the transaction. Thus, the amount of capital used to cover transactions must be of the same order as the revenue realized from the transactions. In the case of a large seller, this requirement is likely to be unworkable, due to the enormous capital requirement and financing costs.

- Consider a transaction in which a buyer unfairly rates a seller as untrustworthy. In the case of a cash bond, the seller loses actual money. In the case of the Trunits model, the seller incurs an opportunity cost (the ability to engage in a certain quantity of business in *this particular marketplace* in the future), but does not lose actual money. The unfair penalty in the case of the cash bond is more severe.

3.3. Why a buyer can trust in the system: The incentive for honesty

As noted by Dellarocas [4], "sellers care about buyer feedback primarily to the extent that they believe it might affect their future profits." Here, we consider the expected profits of the seller when executing transactions over a period of time. The unit of 'time' employed is that required for the completion of a transaction, from the initial sale until the buyer ultimately provides feedback.

At the beginning of each unit of time, the value of goods that can be sold is limited by the number of trunits available. We assume here that the seller engages in transactions of the maximum allowed value, since doing otherwise would incur opportunity costs in terms of both profit, and trunits (if the sale is an honest one). The seller might execute a single large transaction, or split the trunits in order to execute several smaller transactions. Where the same ratio r is used for every transaction, and assuming that the cost of goods to the seller is the same in both cases, these scenarios have equivalent outcomes under our model: the total value of the transactions, the total profit, and the total trunits gained are the same in both cases. For this reason, we simplify the analysis by considering only the case where a single transaction is made in each time period. Under these assumptions, in a period of time consisting of h units, our seller engages in a sequence of h transactions.

Profit on a transaction is the difference between the selling price and costs incurred in selling the item. We express cost, c, as a fraction of selling price. Given that the value of a transaction $V = rT$, profit on a honest transaction $P = (1 - c)\, rT$.

Let T_0 represent the seller's available quantity of trust before the first transaction in a sequence. If all transactions from the first to the i-th are executed honestly, then after the i-th transaction the seller's quantity of trust is:

$$T_i = (1+p)^i T_0 .$$

This quantity of trust available after the i-th transaction will allow the next transaction to be of value $r\,(1 + p)^i\, T_0$. The profit earned from executing transaction $i + 1$ will then be

$$P_{i+1} = (1-c)(1+p)^i rT_0$$

The total profit over a sequence of h honest transactions, beginning with trust T_0, then, is:

$$\begin{aligned} P_{s_h} &= \sum_{i=0}^{h-1} (1-c)(1+p)^i rT_0 \\ &= (1-c)\, rT_0 \sum_{i=0}^{h-1} (1+p)^i \end{aligned}$$

The summation above is a geometric series, so it can be expressed in closed form as:

$$P_{s_h} = (1-c)\, rT_0 \left(\frac{(1+p)^h - 1}{p}\right)$$

Now, consider a seller who has engaged in a transaction with value rT_0. She has two choices: either fulfill the transaction honestly, or cheat the buyer. When cheating, she loses all (T_0) of her trunits and realizes a maximum profit on the transaction of rT_0 (in the case where she fails to supply the purchased good at all). If she is honest, she will gain pT_0 trunits while earning a profit of $(1 - c)rT_0$. While cheating results in a larger immediate profit, honesty may be more profitable in the long run, since the trunits will allow her to engage in additional transactions in the future. Let h be the seller's *horizon,* the number of sales she can foresee making in the future, including the current one. For honesty to be economically advantageous, her total expected profit over her horizon must be greater than that realized by cheating on this first transaction. Setting this inequality, and then solving for h:

$$\begin{aligned} P_{s_h} &> rT_0 \\ (1-c)\, rT_0 \left(\frac{(1+p)^h - 1}{p}\right) &> rT_0 \\ (1-c)\left(\frac{(1+p)^h - 1}{p}\right) &> 1 \\ \frac{(1+p)^h - 1}{p} &> \frac{1}{1-c} \\ (1+p)^h &> \frac{p}{1-c} + 1 \\ h &> \log_{p+1}\left(\frac{p}{1-c} + 1\right) \end{aligned}$$

In the inequality above, note that T_0 disappears—the existence of the incentive for honesty is not dependent on the value of the transaction. Charting this inequality for several values of p yields the graph displayed in Figure 3. Points above the curves indicate combinations of cost ratio c and horizon h for which the honesty incentive exists. It is clear from the chart that, unless sellers have extremely low profit margins, honesty is economically advantageous with even very short horizons. For example where $p = 0.5$ and $c = 0.6$, the minimum such horizon is 2 transactions.

Further, it should be noted that the above analysis assumes no cost to the seller in the event he decides to cheat. This is a very conservative assumption—given possible costs incurred outside the mechanism itself (penalties imposed by the market operator, remedies within the legal system, etc.), the incentive for honesty is likely to be even larger than stated above.

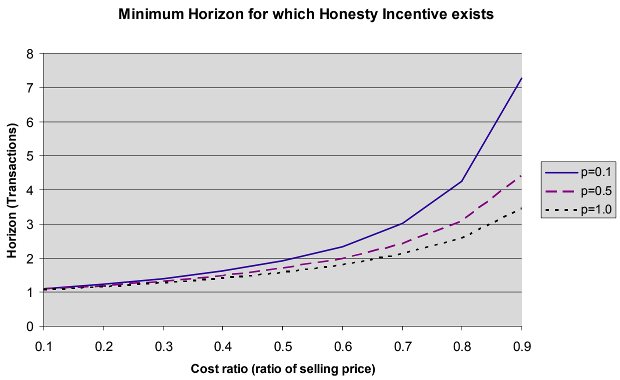

Figure 3: The relationship between cost ratio, horizon, and the honesty incentive

It is unlikely that a seller's cost structure will be known. Since the incentive mechanism is sensitive to the value of c, it is valuable to understand the range of values of c for which the incentive exists. Solving the same inequality for c, we obtain:

$$c \quad < \quad 1 - \frac{p}{(1+p)^h - 1}$$

This inequality yields a chart, shown in Figure 4, which is a reflection of the previous one; points under the curves indicate combinations of c and h for which the incentive exists. It is evident from this graph that the incentive exists even for very high cost ratios. For example, with $p = 0.5$ and a horizon of only three transactions, the incentive exists for cost ratios up to almost 80%.

One potential problem is obvious: if the seller's desired number of future transactions is below this threshold, then the economic incentive for honesty is no longer present. This is an issue common to most trust/reputation systems—it is an instance of the 'exit problem' that was identified above.

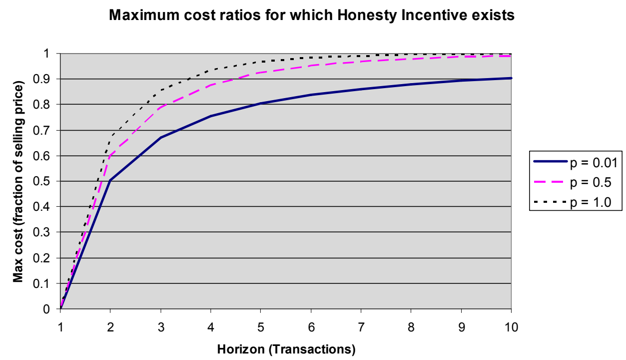

Figure 4: The chart from Figure 3, transformed to highlight ranges of cost ratio where the honesty incentive exists

The Start-up problem

The trunits mechanism, as described to this point, results in a 'chicken-or-egg' problem: you earn trunits only by engaging in honest sales, but you need trunits in order to engage in sales.[2] There is a number of ways in which a seller might enter the market without trunits:

- One possibility is to allow new sellers to put up a cash bond in order to be fronted with a 'loan' of some trunits. Once sufficient trunits had been earned, the seller could repay this loan, reclaiming their bond. This system incurs some of the disadvantages of the cash bond system described above, but such bonds are in use only for short periods of time.
- As a generalization of the above, trunits could be loaned to bonded sellers any time they had an insufficient quantity to engage in a desired transaction. (Note that this addresses the issue of potential restraint of trading volume, described in section 3.2)
- New sellers might gain entry to the market by using a trusted broker. Such brokers exist on eBay today—they maintain high reputations, and sell products on behalf of other people in exchange for a fee. If the brokerage transaction were executed inside the system, then the new seller could gain trunits from the transaction.
- As discussed below, one possible expansion of this system is to adopt a 'free-market' approach, eliminating the required ratio r, and allowing sellers to offer any product/trunit combination they see fit. In this scenario, market forces would determine the value of

[2] While new sellers might simply be provided with an initial quantity of trunits, this opens the door to the re-entry problem, since new users would be in a better position than maximally disreputable ones. Note, however, that this option is worth considering in scenarios where sellers' true identities can be established, and some means external to the mechanism can prevent re-entry.

trunits. Under this system, a new entrant could sell products using 0 trunits, if they were discounted enough to motivate buyers to take the risk. In doing so, sellers could earn trunits for later use.

None of these solutions is clearly superior to the others; more investigation is required.

3.4. The Trunit mechanism's handling of key problems

As discussed above, a system ideally would prevent each of the problems identified in the Vulnerabilities section, without introducing substantial new ones. Thus, we consider how the basic Trunits mechanism deals with each vulnerability.

The 'Reputation Lag problem'

The Trunits mechanism deals directly with this vulnerability by forcing the seller to place trunits in escrow to cover transactions. The Trunits mechanism regulates the rate at which transactions can occur: if the seller holds T trunits, then the maximum value of transactions he can engage in during one unit of time is rT, regardless of timing or circumstances.

The 'Value Imbalance problem'

The Trunits mechanism deals directly with this vulnerability by basing both the quantity of trunits required to cover a transaction, and the size of the reward, on the value of the transaction.

Revisiting the example from the earlier discussion of value imbalance, assume that the seller executes five honest transactions at \$1 each. Using $r = 5$ and $p = 0.2$, $V/r = 5/5 = 1$ trunit required in total to cover the five transactions. Executing these transactions honestly would result in a new trunit balance of $(1 + p)T = (1 + 0.2) * (1) = 1.2$ trunits, or enough to cover a \$6 transaction. The seller is unable to use the reputation gained from the five smaller transactions to cheat a buyer out of \$1,000, as he might in the earlier example.

The 'Ballot-Stuffing problem'

As presented, the Trunits model presents no direct impediment to ballot-stuffing. However, there are many potential means to address this problem; one example follows.

The basic Trunits model uses very simple ratios to determine the risk and reward for transactions. Ballot stuffing is just one reason why more sophisticated functions might be preferable. For example, in many systems (e.g., [5], [14]) only one rating is considered for each buyer, with older ratings being discarded. While we cannot simply discard trunits from previous transactions, we can consider reward functions that drastically reduce the marginal reward from each successive transaction with the same user. This fits with our real world model—recommendations from ten different users are likely to inspire more confidence than a single recommendation from a user who has engaged in ten transactions with a given seller. Similarly, we might reduce the reward realized from transactions with new users, as in [14], in order to deter the use of new accounts for ballot-stuffing.

The 'Re-entry problem'

In [4], Friedman and Resnick's suggestion for dealing with the re-entry problem is discussed: new entrants to the market should incur a cost, such that the cost of re-entry exceeds the benefit. Also in [4], the author cites another of his papers in which he claims to have proved that, in marketplaces with binary feedback mechanisms, the policy of optimal social efficiency is one where new users begin with the worst possible reputation (the same as very disreputable users).

In the Trunits mechanism, new and maximally disreputable users are treated the same, providing no incentive for re-entry, and consistent with the optimal policy. Further, this system is fully compatible with charging fees for new accounts.

The 'Exit problem'

The Trunits mechanism, as described above, is vulnerable to the exit problem. In fact, while the exit problem is common to most trust/reputation mechanisms, it could potentially be magnified under the Trunits mechanism.

Consider a user who has accumulated a large quantity of trust/reputation, and who has decided to leave the market. This user may decide to use her reputation to cheat users before exiting the marketplace.

Under the eBay system, the user might engage in numerous dishonest transactions, but there would be a practical limit on the number that could be completed. Buyers would have some warning—seeing a string of negative transactions in the feedback profile, it would be obvious that the seller's behaviour had deteriorated, and buyers would stop trading with her despite her remaining positive reputation score.

In contrast, under the basic Trunits mechanism the seller would be free to use every available trunit for the purposes of deception, until her supply was exhausted.

While some techniques external to the mechanism might be able to detect such behaviour and stop it at an earlier stage, there is no obvious way to prevent it within the mechanism. However, the risk of this problem may be mitigated by properties of the Trunits model. As has been discussed, trunits have value, in terms of the future profitable transactions they make possible. This value can be determined with some precision, and may be strong incentive for the seller to stay in the market.

While the market operator cannot allow a seller to 'cash in' trunits (for reasons of financial viability), it remains to be established whether it is safe to allow sellers to trade or sell trunits. If the transfer of trunits is prohibited, the seller might honestly profit from their use in at least two ways:

- She might sell her business; the value of the trunits might be considered as an asset in this case, and might be reflected in the selling price of the business.
- She might act as a trusted broker for others who wish to sell products, as discussed above.

A new problem: 'Surplus trust'

The Trunits mechanism, as described here, suffers from another potential problem that may be unique to this system: that of 'surplus trust'. A seller may accumulate trunits beyond what is required to cover his regular transactions.

For example, consider a scenario in which a seller has a fixed production capacity, so he can only sell a fixed quantity per week. Assume that the seller has enough trunits to cover sales at this rate. For each honest sale, the seller receives his own trunits back, plus the reward for honesty. However, the returned trunits are all that is required to cover his future sales. The seller may now use the surplus trunits received as reward to cheat buyers; as long as he does not spend his original trunits on dishonest transactions, he will continue to be able to sell his entire production without impediment.

One possible approach to dealing with this problem is related to a common technique used in other models. Many systems (e.g., [13], [14]) emphasize the most recent rankings by either discarding older rankings, or applying some form of 'decay factor' to de-emphasize them. We might employ a similar idea, with the goal of paring the number of surplus trunits. For example, trunits that have gone unused over a period of time (i.e., the lowest balance in the account over that period of time) are likely to be surplus. A portion of these trunits could be 'taxed'. In a way, this policy has the effect of emphasizing recent history—your level of trustworthiness (i.e., your trunit balance) is tied to your recent activity, even if you were 'more trustworthy' in the past. On a more practical level, the goal of the system is to ensure that, if you are sufficiently trustworthy, you have enough trunits to conduct your business; surplus trunits are, by definition, in excess of the trunits required to conduct business, and paring them is not unreasonable.

4. Future work

Our model is in its early stages, and there are many areas for further investigation, some of which have been mentioned in the preceding sections. Promising areas for future exploration include:

- Addressing potentially untrustworthy buyers, who may rate sellers unfairly. While the impediments to ballot-stuffing discussed above also apply to 'bad-mouthing', other techniques may be useful to detect or discourage unfair feedback.
- Incorporating trading strategy for buyers and sellers. For example, selling agents may still learn to maximize profitability, as in [11], [12]. Buying agents, freed from the need to select 'the most trustworthy seller', might have more flexibility to choose product attributes that meet their preferences.
- Determining how to choose values for market parameters, and studying the possibility of varying the values or r and p. It would be useful to explore marketplaces where not all products require the same degree of trust, and different types of transactions allow different amounts of reward for truthfulness.
- Investigating a free market for trust. While our basic model makes use of a risk ratio that is determined by the market operator, it is worthwhile to explore allowing the market to determine the value of trust. This might allow buyers to assume a greater level of risk in order to obtain better price, or allow sellers to make stronger offers, risking more.
- Allowing anonymity in the marketplace. Our model might allow for marketplaces in which traders are anonymous to one another, yet can still trust one another. While anonymity may allow for a more open marketplace, we would need to ensure that there are not additional challenges for buyers in this kind of scenario.
- Applying the model to contexts other than e-commerce. It would be valuable to transport the Trunits model to more general applications where agents in a multiagent system need to trust each other in order to collaborate.
- Decentralization of the mechanism. The most direct implementation of our model makes use of a centralized agent that provides the account, escrow, and reward services. It is worthwhile to investigate the potential of the model for decentralization.
- Verifying the underlying 'safety' of our proposed approach, to more carefully examine whether any new vulnerabilities are introduced by employing the model in electronic marketplaces.

5. Conclusion

In this paper, we have presented the Trunits model and proposed it as the basis for modeling trust in electronic marketplaces. We have discussed the inherent value of this mechanism, by demonstrating that sellers operating in a Trunits environment will be motivated to be honest under most circumstances. We have shown that the incentive for honesty exists under a broad range of market parameters, even when sellers have very short horizons in terms of intended future transactions. We have also discussed the value of the Trunits approach in addressing many of the challenges that arise when modeling trust in electronic marketplaces, due to the possibility of cheating by trading agents.

One attraction of the Trunits model is the fact that it is very simple computationally, particularly for buyers, who do not need to perform any calculations to determine a seller's trustworthiness. In addition, in its basic form the

Trunits model has minimal storage requirements for the market operator, since only a single trunit balance must be kept for each seller, and a single record for each active transaction. Another attractive feature of the system is the fairness with which it treats sellers—sellers offering the same product are seen as equally trustworthy if they have secured their sale with the required number of trunits.

The Trunits mechanism also facilitates measuring the impact of trust on the profitability of sellers. In many other systems where trust of trading agents is modeled, it is clear that trust has an impact on profitability (since being trustworthy allows more products to be sold and/or fetches higher prices for products), but it is less obvious how to measure that impact. This complicates many issues, including both strategic decisions by sellers, and buyers' interpretations of sellers' feedback. In contrast, the Trunits model offers a precise determination of the value of trust.

As outlined in this paper, there are several promising directions for future work, to continue to refine the model. We plan as well to conduct further validation of the model, to emphasize its value towards the design of truly effective electronic marketplaces.

6. References

[1] Bhattacharjee, R. and Goel, A. 2005. Avoiding ballot stuffing in eBay-like reputation systems. In *Proceeding of the 2005 ACM SIGCOMM Workshop on Economics of Peer-To-Peer Systems* (Philadelphia, Pennsylvania, USA, August 22 - 22, 2005). P2PECON '05. ACM Press, New York, NY, 133-137.

[2] Dellarocas, C. 2001. Analyzing the economic efficiency of eBay-like online reputation reporting mechanisms. In *Proceedings of the 3rd ACM Conference on Electronic Commerce* (Tampa, Florida, USA, October 14 - 17, 2001). EC '01. ACM Press, New York, NY, 171-179.

[3] Dellarocas, C. 2002. Goodwill Hunting: An Economically Effi-cient Online Feedback Mechanism for Environments with Vari-able Product Quality, Lecture Notes in Computer Science, Vol-ume 2531, 238 – 252.

[4] Dellarocas, C. 2003. The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Manage. Sci.* 49, 10 (Oct. 2003), 1407-1424.

[5] eBay. http://pages.ebay.com/help/feedback

[6] Griffiths, N. 2005. Task delegation using experience-based multi-dimensional trust. In *Proceedings of the Fourth international Joint Conference on Autonomous Agents and Multiagent Systems* (The Netherlands, July 25 - 29, 2005). AAMAS '05. ACM Press, New York, NY, 489-496.

[7] Gupta, M., Judge, P., and Ammar, M. 2003. A reputation system for peer-to-peer networks. In *Proceedings of the 13th international Workshop on Network and Operating Systems Support For Digital Audio and Video* (Monterey, CA, USA, June 01 - 03, 2003). NOSSDAV '03. ACM Press, New York, NY, 144-152.

[8] Marsh, S. *Formalising Trust as a Computational Concept.* PhD thesis. University of Stirling. 1994.

[9] Sabater, J. and Sierra, C. 2001. REGRET: reputation in gregarious societies. In *Proceedings of the Fifth international Conference on Autonomous Agents* (Montreal, Quebec, Canada). AGENTS '01. ACM Press, New York, NY, 194-195.

[10] Sabater, J. and Sierra, C. 2005. Review on Computational Trust and Reputation Models. *Artif. Intell. Rev.* 24, 1 (Sep. 2005), 33-60.

[11] Tran, T. and Cohen, R. 2002. A Learning Algorithm for Buying and Selling Agents in Electronic Marketplaces. In *Proceedings of the 15th Conference of the Canadian Society For Computational Studies of intelligence on Advances in Artificial intelligence* (May 27 - 29, 2002). R. Cohen and B. Spencer, Eds. Lecture Notes In Computer Science, vol. 2338. Springer-Verlag, London, 31-43.

[12] Tran, T. and Cohen, R. 2004. Improving User Satisfaction in Agent-Based Electronic Marketplaces by Reputation Modelling and Adjustable Product Quality. In *Proceedings of the Third international Joint Conference on Autonomous Agents and Multiagent Systems - Volume 2* (New York, New York, July 19 - 23, 2004). International Conference on Autonomous Agents. IEEE Computer Society, Washington, DC, 828-835.

[13] Yu, B. and Singh, M.P. 2002. Distributed Reputation Management for Electronic Commerce. In *Computational Intelligence*, 18(4): 535-549, 2002.

[14] Zacharia, G., Moukas, A., and Maes, P. 1999. Collaborative Reputation Mechanisms in Electronic Marketplaces. In *Proceedings of the Thirty-Second Annual Hawaii international Conference on System Sciences-Volume 8 - Volume 8* (January 05 - 08, 1999). HICSS. IEEE Computer Society, Washington, DC, 8026.