# The Beta Reputation System

**Audun Jøsang**

Distributed Systems Technology Centre [*]
Queensland University of Technology, GPO Box 2434, Brisbane Qld 4001, Australia
tel:+61-7-3864 1051, fax:+61-7-3864 1282
email: ajosang@dstc.edu.au


**Roslan Ismail**

Information Security Research Centre
Queensland University of Technology, GPO Box 2434, Brisbane Qld 4001, Australia
tel:+61-7-3864 2575, fax:+61-7-3221 2384
email: r.ismail@student.qut.edu.au

**Abstract**

*Reputation systems can be used to foster good behaviour and to encourage adherence to contracts in e-commerce. Several reputation systems have been deployed in practical applications or proposed in the literature. This paper describes a new system called the beta reputation system which is based on using beta probability density functions to combine feedback and derive reputation ratings. The advantage of the beta reputation system is flexibility and simplicity as well as its foundation on the theory of statistics.*

## 1 Introduction

Contracts and agreements need some form of enforcement in order to be respected. Traditionally, transaction parties can rely on legal procedures in case of disagreement or contract breach. In e-commerce it can be difficult to rely on legal procedures because it is often unclear which jurisdiction applies, and because the cost of legal procedures often are higher than the contractual value itself.

As a substitute for enforcement principles that are used in traditional commerce, reputation systems have emerged as a method for stimulating adherence to electronic contracts and for fostering trust amongst strangers in e-commerce transactions [9]. A reputation system gathers, distributes, and aggregates feedback about participants behaviour. According to Resnick *et al.* [10] reputation mechanisms can provide an incentive for honest behaviour and help people make decisions about who to trust.

Past experience with remote transaction partners can be projected into the future, giving a measure of their trustworthiness. This effect has been called the 'shadow of the future' by political scientist Robert Axelrod [1]. Without such systems, where strangers are interacting in an e-commerce setting, the temptation to act deceptively for immediate gain could be more appealing than cooperation.

The first Web sites to introduce reputation schemes were on-line auction sites such as eBay.com. They are now also used by company reputation rating sites such as BizRate.com, which ranks merchants on the basis of customer ratings. Consumer Reports Online's eRatings, rates merchants on the basis of test purchases carried out by Consumer Reports staff. Product review sites have also emerged, such as Epinions.com, in which reviews themselves are actually rated by other reviewers. Except for eRatings, most of the systems do little to overcome the issue of establishing initial trust for new merchants in the e-commerce arena, as strong reputation ratings generally require time to develop.

In the physical world, capturing and distributing feedback can be costly. In comparison, the Internet is extremely efficient. However reputation systems still encounter significant challenges. Feedback can get erased if an entity changes its name, and a dishonest participant can use this to start fresh every time it builds up a bad reputation. People may not bother to provide feedback at all, negative feedback can be difficult to elicit, and it is difficult to ensure that feedback is honest [10]. One example of dishonesty through reputation systems is the attempt by three men to sell a fake painting on eBay for $US135,805 [13]. The sale was abandoned just prior to purchase when the buyer became suspicious. It was shown that two of the fraudsters actually had good Feedback Forum ratings, developed through rating each other favourably, and by engaging in honest sales prior to the fraudulent attempt. Reputation can thus be seen as an asset, not only to promote oneself, but also as something that can be cashed in through a fraudulent transaction with high gain. It emerges that reputation systems have a multitude of complex facets, and is becoming a fertile ground for research[1]. Two fundamental aspects to consider are:

1. An *engine* that calculates the value of the users' reputation ratings from various inputs including feedback from other users. Many specific reputation engines have been proposed recently in the literature. These range from simple ones which accept numerical values which are simply added together, such as in eBay, to advanced mechanisms which deploy more complex mathematical equations [15, 14, 11, 3, 7, 8].

2. A *propagation mechanism* that allows entities to obtain reputation values when required. There are two available approaches for user reputation propagation. In the *centralised approach* reputation values are stored in a central server, and whenever there is a need, users forward their query to the central server for the reputation value. One of practical example is eBay. In the *decentralised* approach everybody keeps and manages reputation of other people themselves. Whenever there is a need, users can ask others for the required reputation values [16, 5, 14].

Up to now, most of the researches have concentrated on developing of reputation engine, and that is also the focus of our paper. However, we would like to point out that development of robust and secure mechanisms for reputation propagation is a topic that deserves much more study.

In this paper we propose a new reputation engine called the *beta reputation system* which is based on the beta probability density function. In contrast to most other reputation systems which are intuitive and ad hoc, the beta reputation system has a firm basis in the theory of statistics. Although we describe a centralised approach, the beta reputation system can also be used in a distributed setting.

---

[1]The Reputations Research Network (http://databases.si.umich.edu/reputations/) is a forum for people conducting research into how reputation systems should work in theory, how they actually work in practice, and how they could work better.

## 2 Building Blocks in the Beta Reputation System

The beta reputation system consists of elements that can be used separately or in combination in order to provide a flexible framework for integrating reputation services into e-commerce applications. The reputation function and reputation rating which are described in Sections 2.2 and 2.3 below form a basis on which the building blocks described in the subsequent sections can be added depending on the requirements.

### 2.1 The Beta Density Function

Our reputation system is based on the beta probability density function which can be used to represent probability distributions of binary events. This provides a sound mathematical basis for combining feedback and for expressing reputation ratings. The mathematical analysis leading to the expression for posteriori probability estimates of binary events can be found in many text books on probability theory, e.g. Casella & Berger 1990[2] p.298, and we will only present the results here.

Posteriori probabilities of binary events can be represented as beta distributions. The beta-family of probability density functions is a continuous family of functions indexed by the two parameters $\alpha$ and $\beta$. The beta distribution $f(p \mid \alpha, \beta)$ can be expressed using the gamma function $\Gamma$ as:

$$f(p \mid \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} , \quad \text{where } 0 \leq p \leq 1, \ \alpha > 0, \ \beta > 0 , \tag{1}$$

with the restriction that the probability variable $p \neq 0$ if $\alpha < 1$, and $p \neq 1$ if $\beta < 1$. The probability expectation value of the beta distribution is given by:

$$\mathrm{E}(p) = \alpha/(\alpha + \beta). \tag{2}$$

Let us consider a process with two possible outcomes $\{x, \overline{x}\}$, and let $r$ be the observed number of outcome $x$ and let $s$ be the observed number of outcome $\overline{x}$. Then the probability density function of observing outcome $x$ in the future can be expressed as a function of past observations by setting:

$$\alpha = r + 1 \quad \text{and} \quad \beta = s + 1, \quad \text{where} \quad r, s \geq 0 . \tag{3}$$

As an example, a process with two possible outcomes $\{x, \overline{x}\}$ that has produced outcome $x$ seven times and outcome $\overline{x}$ only once, will have a beta function expressed as $f(p \mid 8, 2)$ which is plotted in Figure 1.
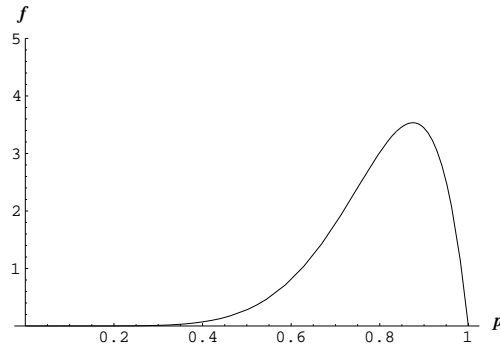


Figure 1: Beta function of event $x$ after 7 observations of $x$ and 1 observation of $\overline{x}$.

This curve expresses the uncertain probability that the process will produce outcome $x$ during future observations. The probability expectation value is given by $E(p) = 0.8$. This can be interpreted as saying that the relative frequency of outcome $x$ in the future is somewhat uncertain, and that the most likely value is 0.8.

The variable $p$ is a probability variable, so that for a given $p$ the probability density $f(p \mid \alpha, \beta)$ represents second order probability. The first-order variable $p$ represents the probability of an event, whereas the density $f(p \mid \alpha, \beta)$ represents the probability that the first-order variable has a specific value. Since the first-order variable $p$ is continuous, the second-order probability $f(p \mid \alpha, \beta)$ for any given value of $p \in [0, 1]$ is vanishingly small and therefore meaningless as such. It is only meaningful to compute $\int_{p_1}^{p_2} f(p \mid \alpha, \beta)$ for a given interval $[p_1, p_2]$, or simply to compute the expectation value of $p$. Below we will define a reputation rating that is based on the expectation value.

## 2.2 The Reputation Function

When observing binary processes with two possible outcomes $\{x, \overline{x}\}$, the beta function takes the integer number of past observations of $x$ and $\overline{x}$ to estimate the probability of $x$, or in other words, to predict the expected relative frequency with which $x$ will happen in the future.

By replacing the parameters $(\alpha, \beta)$ in Eq.(1) by $(r, s)$ through the mapping of Eq.(3), the parameters $(r, s)$ can be directly interpreted as the number of observations of outcome $x$ and $\overline{x}$ respectively, and the prior density function before any observation can be expressed by setting $r, s = 0$.

Combining feedback resulting from an e-commerce transaction is not the same as statistical observations of a binary event, because an agent's perceived satisfaction after a transaction is not binary. Instead will let positive and negative feedback be given as a pair $(r, s)$ of continuous values where $r$ reflects the degree of satisfaction and $s$ reflects the degree of dissatisfaction. This leads to the following definition of the reputation function:

**Definition 1 (Reputation Function)** *Let $r_T^X$ and $s_T^X$ respectively represent the (collective) amount of positive and negative feedback about target entity $T$ provided by an agent (or collection of agents) denoted by $X$, then the function $\varphi(p \mid r_T^X, s_T^X)$ defined by:*

$$\varphi(p \mid r_T^X, s_T^X) = \frac{\Gamma(r_T^X + s_T^X + 2)}{\Gamma(r_T^X + 1)\Gamma(s_T^X + 1)} p^{r_T^X} (1 - p)^{s_T^X} , \quad where\ 0 \le p \le 1,\ 0 \le r_T^X,\ 0 \le s_T^X . \quad (4)$$

*is called $T$'s reputation function by $X$. The tuple $(r_T^X, s_T^X)$ will be called $T$'s reputation parameters by $X$. For simplicity and compactness of notation we will sometimes write $\varphi_T^X$ instead of $\varphi(p \mid r_T^X, s_T^X)$.*

By using Eq.(2) the probability expectation value of the reputation function can be expressed as:

$$E(\ \varphi(p \mid r_T^X, s_T^X)\ ) = (r_T^X + 1)/(r_T^X + s_T^X + 2) . \quad (5)$$

We will consider reputation functions to be subjective in the sense that if agent $X$ provides feedback about target $T$, then the reputation function resulting from that feedback represents $T$'s reputation as seen by $X$, and can not be considered to represent $T$'s reputation from an objective viewpoint, because no such thing exists. For this reason $\varphi(p \mid r_T^X, s_T^X)$ is called $T$'s reputation function by $X$. Superscripts thus represents the feedback provider, and subscripts represent the feedback target.

## 2.3 The Reputation Rating

The concept of reputation function is ideal for mathematical manipulation, but less so for communicating a reputation rating to human users. A more simple representation than the reputation function is therefore needed. Most people are familiar with the notion of a probability value, and the probability expectation value $\mathrm{E}(p)$ from Eq.(5) therefore seems very suitable. This would give a reputation rating in the range $[0, 1]$ where the value $0.5$ represents a neutral rating.

It might be more intuitive to let the reputation rating be in a range such as e.g. $[-100, +100]$ or $[-1, +1]$ where the zero middle value represents neutral rating. Any range can be obtained by simply scaling the probability expectation value. In the following we will scale the reputation rating to be in the range $[-1, +1]$.

**Definition 2 (Reputation Rating)** *Let $r_T^X$ and $s_T^X$ respectively represent the (collective) amount of positive and negative feedback about target entity $T$ provided by an agent (or collection of agents) denoted by $X$, then the function $\mathrm{Rep}(r_T^X, s_T^X)$ defined by:*

$$\mathrm{Rep}(r_T^X, s_T^X) \;=\; \left(\mathrm{E}(\,\varphi(p\,|\,r_T^X, s_T^X)\,) - 0.5\right) \cdot 2 \;=\; \frac{r_T^X - s_T^X}{r_T^X + s_T^X + 2}\,. \tag{6}$$

*is called $T$'s reputation rating by $X$. For simplicity and compactness of notation we will sometimes write $\mathrm{Rep}_T^X$ instead of $\mathrm{Rep}(r_T^X, s_T^X)$.*

The reputation rating can be interpreted as measure of reputation, or in other words as an indication of how a particular agent is expected to behave in future transactions. The parameter $r^X$ represents the amount of feedback about $T$ supporting good reputation, and the parameters $s^X$ represents the amount of feedback about $T$ supporting bad reputation.

## 2.4 Combining Feedback

Reputation systems must be able to combine feedback from multiple sources. In the beta reputation system this can be done by simply accumulating all the received $r$ and $s$ parameters from the feedback providers. Assume two agents $X$ and $Y$ giving feedback about the same target $T$. Target $T$'s reputation functions by $X$ and by $Y$ can then be expressed as $\varphi(p\,|\,r_T^X, s_T^X)$ and $\varphi(p\,|\,r_T^Y, s_T^Y)$ respectively, or as $\varphi_T^X$ and $\varphi_T^Y$ in the short notation. By combining their feedback, $X$ and $Y$ can create an updated reputation function for $T$. The following definition captures this idea.

**Definition 3 (Combining Feedback)**
*Let $\varphi(p\,|\,r_T^X, s_T^X)$ and $\varphi(p\,|\,r_T^Y, s_T^Y)$ be two different reputation functions on $T$ resulting from $X$ and $Y$'s feedback respectively. The reputation function $\varphi(p\,|\,r_T^{X,Y}, s_T^{X,Y})$ defined by:*

$$\begin{aligned} &1. \quad r_T^{X,Y} = r_T^X + r_T^Y \\ &2. \quad s_T^{X,Y} = s_T^X + s_T^Y \end{aligned} \tag{7}$$

*is then called $T$'s combined reputation function by $X$ and $Y$. By using the symbol '$\oplus$' to designate this operator, we get $\varphi(p\,|\,r_T^{X,Y}, s_T^{X,Y}) = \varphi(p\,|\,r_T^X, s_T^X) \oplus \varphi(p\,|\,r_T^Y, s_T^Y)$. In the short notation this can be written as: $\varphi_T^{X,Y} = \varphi_T^X \oplus \varphi_T^Y$.*

It is easy to prove that $\oplus$ is both commutative and associative which means that the order in which reputation functions are combined has no importance. Independence between reputation functions must be assumed, which for example translates into not allowing the same reputation function to be counted more than once.

## 2.5  Discounting

Intuitively, feedback from highly reputed agents should carry more weight than feedback from agents with low reputation rating. This can be taken into account by discounting the feedback as a function of the reputation of the agent who provided the feedback. This type of discounting has been described in the context of belief theory by Shafer 1976 [12] and by Jøsang [6].

Jøsang's belief model uses a metric called *opinion* to describe beliefs about the truth of statements. An opinion is a tuple $\omega_x^A = (b, d, u)$, where the parameters $b$, $d$ and $u$ represent *belief*, *disbelief* and *uncertainty* parameters respectively. These parameters satisfy $b + d + u = 1$ where $b, d, u \in [0, 1]$. This is a sub-additive probability model where the probability of truth ($b$) and the probability of falsehood ($d$) not necessarily add up to 1. The $b$ parameter can thus be interpreted as the probability that proposition $x$ is true. The $d$ parameter can be interpreted as the probability that $x$ is false. Finally, the $u$ parameter represents the probability mass that is unaccounted for; in effect, by assigning probability mass to $u$ the subject proclaims his inability to assess the probability value of $x$. The advantage of sub-additivity is that it is possible to express degrees of uncertainty regarding the probability of a particular event. The following definition is taken from Jøsang [6].

**Definition 4 (Belief Discounting)**
*Let $X$ and $Y$ be two agents where $\omega_Y^X = (b_Y^X, d_Y^X, u_Y^X)$ is $X$'s opinion about $Y$'s advice, and let $T$ be the target agent where $\omega_T^Y = (b_T^Y, d_T^Y, u_T^Y)$ is $Y$'s opinion about $T$ expressed in an advice to $X$. Let $\omega_T^{X:Y} = (b_T^{X:Y}, d_T^{X:Y}, u_T^{X:Y})$ be the opinion such that:*

$$
\begin{aligned}
&\textbf{1.} \quad b_T^{X:Y} = b_Y^X b_T^Y \;, \\
&\textbf{2.} \quad d_T^{X:Y} = b_Y^X d_T^Y \;, \\
&\textbf{3.} \quad u_T^{X:Y} = d_Y^X + u_Y^X + b_Y^X u_T^Y \;,
\end{aligned}
\tag{8}
$$

*then $\omega_T^{X:Y}$ is called the discounting of $\omega_T^Y$ by $\omega_Y^X$ expressing $X$'s opinion about $T$ as a result of $Y$'s advice to $X$. By using the symbol '$\otimes$' to designate this operator, we can write $\omega_T^{X:Y} = \omega_Y^X \otimes \omega_T^Y$.*

The opinion metric can be interpreted equivalently to the beta function, and Jøsang [6] provides a mapping between the two representations defined by:

$$
\begin{cases}
b = \frac{r}{r+s+2} \;, \\
d = \frac{s}{r+s+2} \;, \\
u = \frac{2}{r+s+2} \;.
\end{cases}
\tag{9}
$$

By using Eq.(9) to insert the expressions for $(b, d, u)$ in Eq.(8), we obtain the following definition of the discounting operator for reputation functions.

**Definition 5 (Reputation Discounting)**
*Let $X$, $Y$ and $T$ be three agents where $\varphi(p \,|\, r_Y^X, s_Y^X)$ is $Y$'s reputation function by $X$, and $\varphi(p \,|\, r_T^Y, s_T^Y)$ is $T$'s reputation by $Y$. Let $\varphi(p \,|\, r_T^{X:Y}, s_T^{X:Y})$ be the reputation function such that:*

$$
\textbf{1.} \quad r_T^{X:Y} = \frac{2 r_Y^X r_T^Y}{(s_Y^X + 2)(r_T^Y + s_T^Y + 2) + 2 r_Y^X} \;,
$$

$$
\textbf{2.} \quad s_T^{X:Y} = \frac{2 r_Y^X s_T^Y}{(s_Y^X + 2)(r_T^Y + s_T^Y + 2) + 2 r_Y^X} \;,
\tag{10}
$$

*then it is called $T$'s discounted reputation function by $X$ through $Y$. By using the symbol '$\otimes$' to designate this operator, we can write $\varphi(p \,|\, r_T^{X:Y}, s_T^{X:Y}) = \varphi(p \,|\, r_Y^X, s_Y^X) \otimes \varphi(p \,|\, r_T^Y, s_T^Y)$. In the short notation this can be written as: $\varphi_T^{X:Y} = \varphi_Y^X \otimes \varphi_T^Y$.*

It is easy to prove that $\otimes$ is associative but not commutative. This means that in case of a chain of reputation functions, the computation can start in either end of the chain, but that the order of the reputation functions is significant. In a chain including more than two reputation function, independence must be assumed, which for example translates into not allowing the same reputation function to appear more than once.

## 2.6  Forgetting

Old feedback may not always be relevant for the actual reputation rating, because the agent may change its behaviour over time. What is needed is a model in which old feedback is given less weight than more recent feedback. This translates into gradually forgetting old feedback. This can be achieved by introducing a forgetting factor which can be adjusted according to the expected rapidity of change in the observed entity. We will use a forgetting scheme first described by Jøsang [4].

Assume a collection of agents who have provided a sequence $Q$ containing $n$ feedback tuples $(r_{T,i}^Q, s_{T,i}^Q)$ indexed by $i$ about target $T$. The combined feedback from all feedback tuples in $Q$ can be expressed as:

$$r_T^Q = \sum_{i=1}^n r_{T,i}^Q \quad \text{and} \quad s_T^Q = \sum_{i=1}^n s_{T,i}^Q \; . \tag{11}$$

The forgetting factor can now be introduce so that:

$$r_{T,\lambda}^Q = \sum_{i=1}^n r_{T,i}^Q \, \lambda^{(n-i)} \quad \text{and} \quad s_{T,\lambda}^Q = \sum_{i=1}^n s_{T,i}^Q \, \lambda^{(n-i)} \; , \quad \text{where } 0 \le \lambda \le 1. \tag{12}$$

It can be observed that having $\lambda = 1$ is equivalent to not having a forgetting factor, i.e. that nothing is forgotten. The other extreme is when $\lambda = 0$ resulting in only the last feedback value to be counted and all others to be completely forgotten. It is important to observe that the order in which feedback was received plays a key role.

One disadvantage of Eq.(12) is that all feedback tuples in the sequence $Q$ must be kept forever. This can be avoided by deriving a recursive algorithm for computing the $(r, s)$ parameters. We define the recursive parameters $(r_{T,\lambda}^{Q(i)}, s_{T,\lambda}^{Q(i)})$ so that:

$$r_{T,\lambda}^{Q(i)} = r_{T,\lambda}^{Q(i-1)} \lambda + r_{T,i}^Q \quad \text{and} \quad s_{T,\lambda}^{Q(i)} = s_{T,\lambda}^{Q(i-1)} \lambda + s_{T,i}^Q \; , \quad \text{where } 0 \le \lambda \le 1. \tag{13}$$

The $(r_{T,\lambda}^Q, s_{T,\lambda}^Q)$ parameters resulting from the sequence $Q$ containing $n$ feedback values, with a forgetting factor $\lambda$ can then be expressed as:

$$r_{T,\lambda}^Q = r_{T,\lambda}^{Q(n)} \quad \text{and} \quad s_{T,\lambda}^Q = s_{T,\lambda}^{Q(n)} \; , \quad \text{where } 0 \le \lambda \le 1. \tag{14}$$

## 2.7  Providing and Collecting Feedback

After each transaction, a single agent can provide both positive and negative feedback simultaneously, in the form of the parameters $r \ge 0$ and $s \ge 0$ respectively. The purpose of providing feedback as a pair $(r, s)$ is to reflect the idea that an agent's performance in a transaction for example can be partly satisfactory. This could be expressed through a feedback with $r, s = 0.5$ which carries the weight of 1, whereas a feedback with $r, s = 0$ is weight-less, and would be as if no feedback was provided at all. In general the sum $r + s$ can be interpreted as the weight of the feedback. The feedback can thus carry an arbitrary high weight by making the sum $r + s$ arbitrary high, and carries the minimum weight when $r + s = 0$, where the latter case is equivalent to not providing any feedback at all.

As an alternative to providing feedback as a pair $(r, s)$, it is possible to define a normalisation weight denoted by $w$ so that the sum of the $(r, s)$ parameters satisfy $r + s = w$. The normalisation will allow the feedback to be provided as a single value. Similarly to the reputation rating, the single value feedback can be defined on any range such as e.g. $[-100, +100]$ or $[-1, +1]$. In the following the single value feedback value will be defined by $v$, such that $v \in [-1, +1]$. The $(r, s)$ pair can then be derived as a function of $w$ and $v$ according to:

$$r = w(1 + v)/2 \quad \text{and} \quad s = w(1 - v)/2 . \tag{15}$$

The purpose of using $v$ instead of $(r, s)$ is to make it easy for people to provide feedback. The idea is that most people are familiar with a single valued rating, whereas being required to provide feedback as a pair $(r, s)$ could be confusing for many.

The normalisation weight $w$ can be used to let the feedback weight be determined by the the value of the transaction, so that feedback for high value transactions would carry more weight, i.e. the feedback would translate into high $(r, s)$ values. The purpose of this can for example be to make a reputation rating immune against being overly influenced by feedback from transactions with negligible value.

It is assumed that feedback is received and stored by a feedback collection centre which can be denoted by $C$. In case discounting is excluded, the resulting reputation functions can be considered to originate from the centre itself, and the reputation function of target $T$ can then be denoted by $\varphi_T^C$, although the feedback in reality came from other agents. In case discounting is taken into consideration, the reputation function of target $T$ can be denoted by $\varphi_T^{C:Q}$ where $Q$ denotes the body of collected feedback about $T$. The expression $\varphi_T^{C:Q}$ should be interpreted as $T$'s reputation function by $C$ derived from feedback values in $Q$ which all have been discounted by the reputation of the feedback providers.

Figure 2 below shows a typical reputation framework, where it is assumed that all agents, denoted by $X$ and $Y$ in this example, are authenticated, and that no agent can change identity.
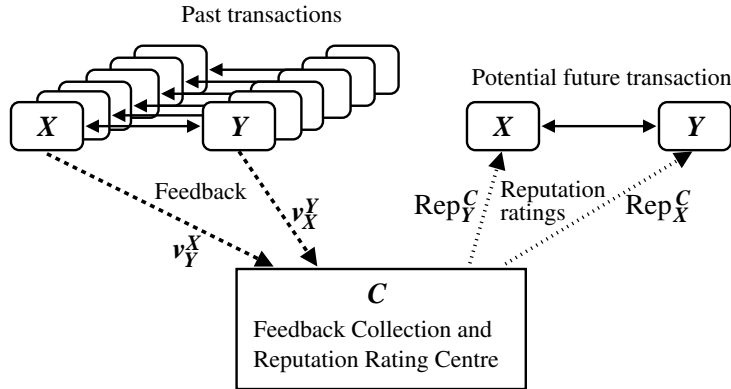


Figure 2: Framework for collecting feedback and providing reputation ratings

After a transaction is completed, the agents provide feedback about each others performance during the transaction. In the figure it can be seen that $X$ provides feedback about $Y$ and that $Y$ provides feedback about $X$. The reputation centre collects feedback from all the agents. The centre discounts received feedback as a function of the feedback providing agent's reputation, before updating the target agent's reputation function and rating. Updated reputation ratings are provided online for all the agents to see, and is used by the agents to decide whether or not to transact with a particular agent. In the figure it can be seen that $X$ and $Y$ fetch each others reputation rating.

# 3 Performance of the Beta Reputation System

This section provides some examples to illustrate how the various building blocs described above influences the reputation rating. Feedback will be provided as a single value $v$ instead of as a pair $(r, s)$.

## 3.1 Example A: Varying Weight

This example shows how the reputation rating evolves as a function of accumulated positive feedback with varying weight $w$. Let the centre $C$ receive a sequence $Q$ of $n$ identical feedback values $v = 1$ about target $T$. Discounting and forgetting are not considered. By using Eqs.(6), (7), and (15), $T$'s reputation parameters and rating can be expressed as a function of $n$ and $w$ according to:

$$\begin{cases} r_T^C & = nw \\ s_T^C & = 0 \end{cases} \qquad \text{Rep}_T^C = \frac{nw}{nw+2} \qquad (16)$$

Figure 3 below shows the reputation rating as a function of $n$ for $w = 0.0$, $w = 0.1$, $w = 0.2$, $w = 0.6$ and $w = 1.0$.
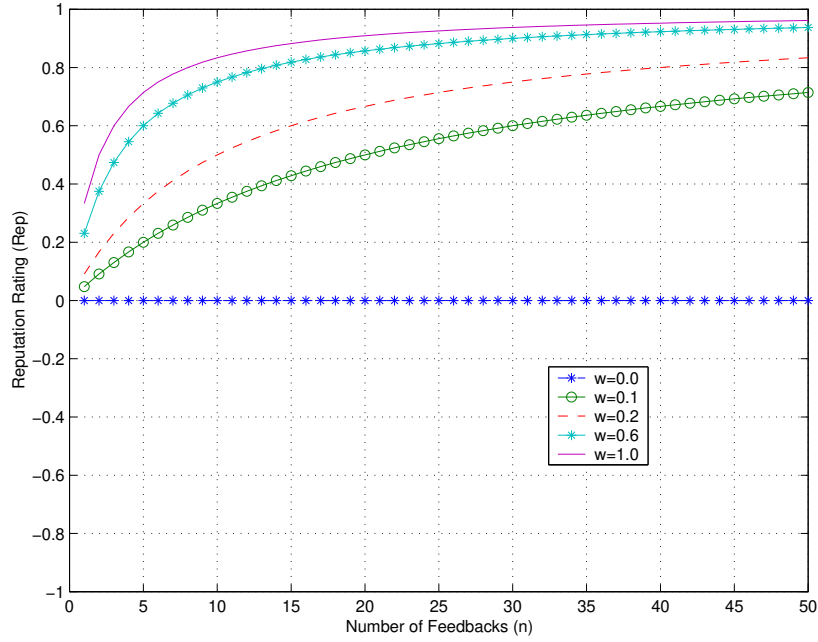


Figure 3: Reputation rating as a function of $n$ with varying weight $w$.

It can be seen that the rating quickly approaches 1 when the $w = 1$ whereas when $w = 0$ the start rating Rep $= 0$ remains unchanged and is thus not influenced by the feedback at all.

## 3.2 Example B: Varying Feedback

This example shows how the reputation rating evolves as a function of accumulated feedback with fixed weight $w = 1$ and varying feedback value $v$. Let the agent $T$ get a sequence $Q$ of $n$ identical feedback values $v$. Discounting and forgetting are not considered. By using Eqs.(6), (7), and (15), $T$'s reputation parameters and rating can be expressed as a function of $n$ and $v$ according to:

$$\begin{cases} r_T & = \frac{n(1+v)}{2} \\[2ex] s_T & = \frac{n(1-v)}{2} \end{cases} \qquad \mathrm{Rep}_T = \frac{nv}{n+2} \qquad (17)$$

Figure 4 below shows the reputation rating as a function of $n$ for $v = 1.0$, $v = 0.5$, $v = 0.0$, $v = -0.5$, $v = -1.0$.
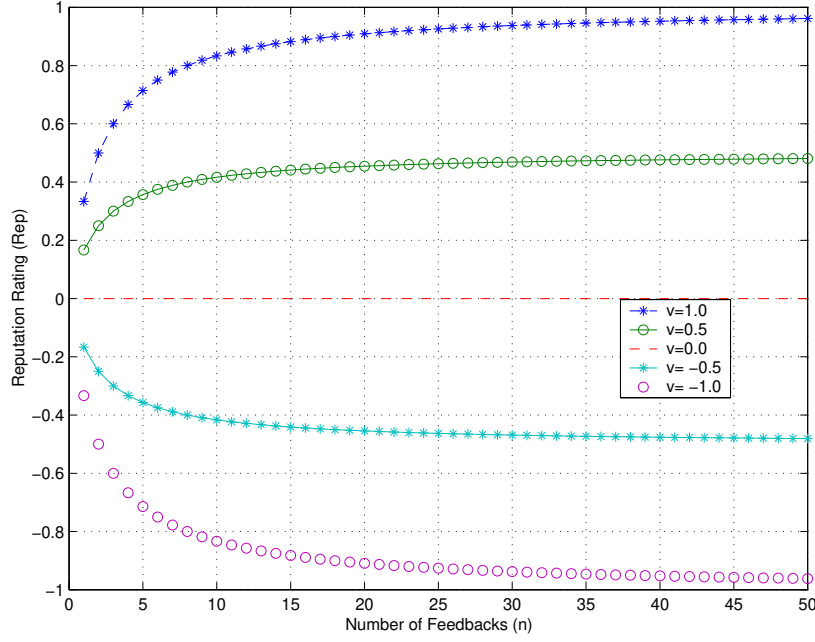


Figure 4: Reputation rating as a function of $n$ with varying feedback $v$.

It can be seen that the rating quickly becomes relatively stable when the feedback value is fixed. For $v = 1$ the rating approaches 1, and for $v = -1$ the rating approaches -1.

### 3.3 Example C: Varying Discounting

This example shows how the reputation rating evolves as a function of accumulated feedback with fixed weight $w$ and varying discounting. Let $w = 1$ and let centre $C$ receive a sequence $Q$ of $n$ identical feedback values $v = 1$ about target $T$. This time discounting is included. Each feedback tuple $(r_T^X, s_T^X) \in Q$ (with fixed value (1,0)) is thus discounted as a function of the feedback provider $X$'s reputation function defined by $(r_X^C, s_X^C)$. Forgetting is not considered. By using Eqs.(6), (7), (10), and (15) , $T$'s reputation parameters and rating can be expressed as a function of $n$ and $(r_X^C, s_X^C)$ according to:

$$\begin{cases} r_T^{C:Q} & = \frac{2nr_X^C}{3(s_X^C+2)+2r_X^C} \\[2ex] s_T^{C:Q} & = 0 \end{cases} \qquad \mathrm{Rep}_T^{C:Q} = \frac{nr_X^C}{nr_X^C+2r_X^C+3s_X^C+6} \qquad (18)$$

Figure 5 below shows the reputation rating as a function of $n$ for $(r_X^C = 1000, s_X^C = 0)$, $(r_X^C = 5, s_X^C = 0)$, $(r_X^C = 3, s_X^C = 0)$, $(r_X^C = 3, s_X^C = 3)$, and $(r_X^C = 0, s_X^C = 0)$.
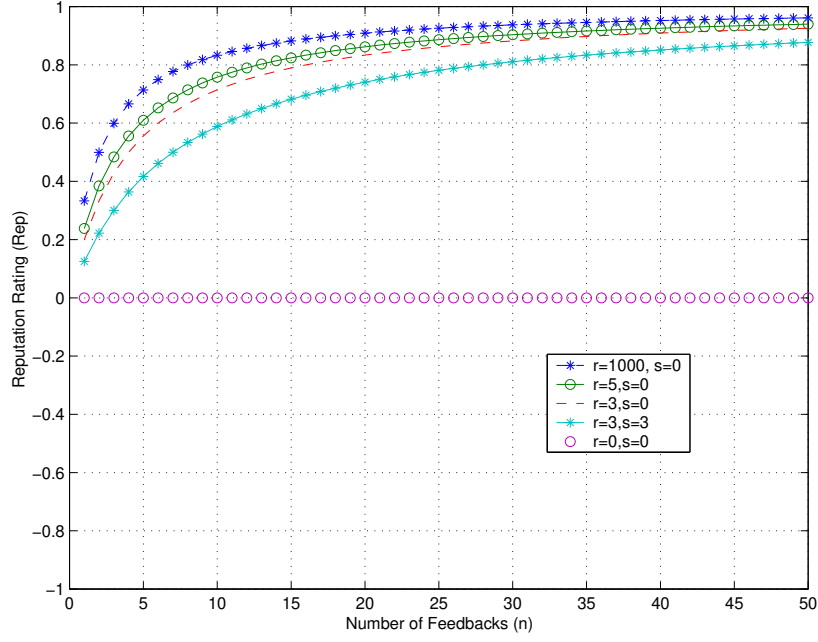
Figure 5: Reputation rating as a function of $n$ with varying discounting rate.

It can be seen that the reputation function discounted by $(r_X^C = 1000, s_X^C = 0)$ is practically equivalent to no discounting at all because the curve is indistinguishable from the uppermost curve in Figure 3 and Figure 4. As $X$'s reputation function gets "weaker" $T$'s rating is less influenced by the feedback from $X$, and when discounted by $(r_X^C = 0, s_X^C = 0)$, $X$'s start rating Rep = 0 remains unchanged and is thus not influenced by the feedback at all.
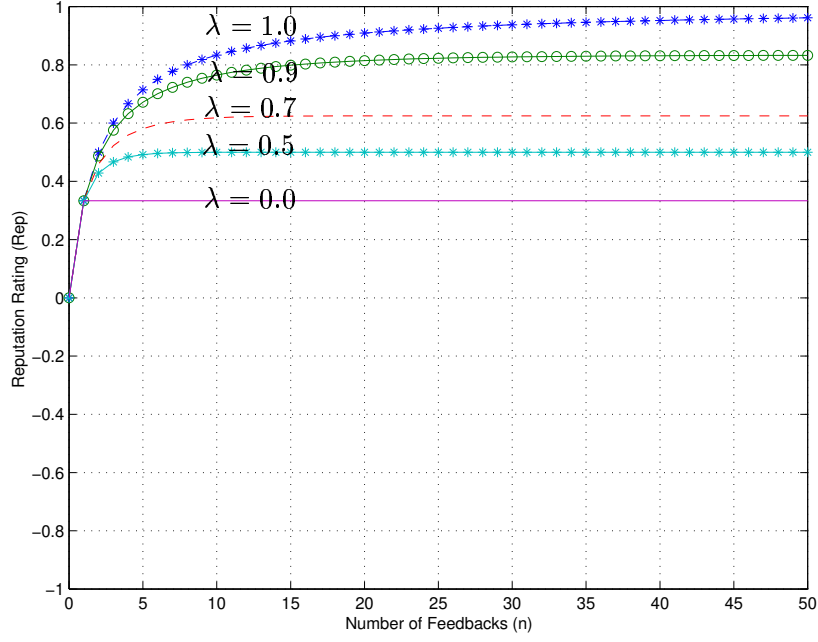
### 3.4 Example D: Varying Forgetting Factor

This example shows how the reputation rating evolves as a function of accumulated feedback with fixed weight $w$ and varying forgetting factor $\lambda$. Let $w = 1$ and let centre $C$ receive a sequence $Q$ of $n$ identical feedback values $v = 1$ about target $T$. Discounting is not considered. By using Eqs.(6), (7), (12), (15) and the fact that $\sum_{i=0}^{n} \lambda^i = (1 - \lambda^{n+1}/(1 - \lambda)$ when $\lambda < 1$, $T$'s reputation parameters and rating can be expressed as a function of $n$ and $\lambda$ according to:

$$
\begin{cases} r_{T,\lambda}^C = \frac{1-\lambda^n}{1-\lambda} \\[2mm] s_{T,\lambda}^C = 0 \end{cases} \quad \mathrm{Rep}_{T,\lambda}^C = \frac{1-\lambda^n}{3-2\lambda-\lambda^n} \qquad \text{when } \lambda \neq 1
$$

$$
\begin{cases} r_{T,\lambda}^C = n \\[2mm] s_{T,\lambda}^C = 0 \end{cases} \quad \mathrm{Rep}_{T,\lambda}^C = \frac{n}{n+2} \qquad \text{when } \lambda = 1
$$

(19)

Figure 6 below shows the reputation rating as a function of $n$ for $\lambda = 1.0$, $\lambda = 0.9$, $\lambda = 0.7$, $\lambda = 0.5$, and $\lambda = 0.0$.

It can be seen that the value of $\lambda$ determines the maximum reputation rate that an agent can get. With $\lambda = 1$, nothing is forgotten, and the maximum rate is 1. With $\lambda = 0$ only the last feedback is remembered, and the maximum reputation rate is $1/3$.

Figure 6: Reputation rating as a function of $n$ with varying forgetting factor $\lambda$.

### 3.5 Example E: Varying Feedback and Varying Forgetting Factor

This example shows how the reputation rating evolves as a function of accumulated feedback with fixed weight $w = 1$. Discounting is not considered. Let there be a sequence $Q$ of 50 feedback inputs about $T$, where the first 25 have value $v_{T,i}^Q = +1$, and the subsequent 25 inputs have value $v_{T,i}^Q = -1$. By using Eqs.(6), (7), (12), and (15), $T$'s reputation parameters and rating can be expressed as a function of $n$, $v$ and $\lambda$ according to:

$$
\begin{cases}
r_{T,\lambda}^Q = \frac{1}{2} \sum_{i=1}^n \left(1 + v_{T,i}^Q\right) \lambda^{(n-i)} \\[2mm]
s_{T,\lambda}^Q = \frac{1}{2} \sum_{i=1}^n \left(1 - v_{T,i}^Q\right) \lambda^{(n-i)}
\end{cases}
\qquad
\mathrm{Rep}_T^{Q(n)} = \frac{\sum_{i=1}^n v_{T,i}^Q \, \lambda^{(n-i)}}{2 + \sum_{i=1}^n \lambda^{(n-i)}}
\tag{20}
$$

In this particular example, Eq.(20) can be expressed in a more explicit form according to:

$$
\begin{cases}
r_{T,\lambda}^C = \frac{1-\lambda^n}{1-\lambda} \\[2mm]
s_{T,\lambda}^C = 0
\end{cases}
\qquad
\mathrm{Rep}_{T,\lambda}^C = \frac{1-\lambda^n}{3-2\lambda-\lambda^n}
\qquad
\text{when}
\quad
\begin{cases}
n \in [0, 25] \\
\lambda \neq 1
\end{cases}
$$

$$
\begin{cases}
r_{T,\lambda}^C = \frac{1-\lambda^{25}}{1-\lambda} \cdot \lambda^{(n-25)} \\[2mm]
s_{T,\lambda}^C = \frac{1-\lambda^{(n-25)}}{1-\lambda}
\end{cases}
\qquad
\mathrm{Rep}_{T,\lambda}^C = \frac{2\lambda^{(n-25)}-\lambda^n-1}{3-2\lambda-\lambda^n}
\qquad
\text{when}
\quad
\begin{cases}
n \in [26, 50] \\
\lambda \neq 1
\end{cases}
\tag{21}
$$

$$
\begin{cases} r_{T,\lambda}^{C} = n \\ \\ s_{T,\lambda}^{C} = 0 \end{cases} \qquad \mathrm{Rep}_{T,\lambda}^{C} = \frac{n}{n+2} \qquad \text{when} \quad \begin{cases} n \in [0, 25] \\ \lambda = 1 \end{cases}
$$

(22)

$$
\begin{cases} r_{T,\lambda}^{C} = 25 \\ \\ s_{T,\lambda}^{C} = n - 25 \end{cases} \qquad \mathrm{Rep}_{T,\lambda}^{C} = \frac{50-n}{n+2} \qquad \text{when} \quad \begin{cases} n \in [26, 50] \\ \lambda = 1 \end{cases}
$$

Figure 7 below shows the reputation rating as a function of $Q$ and $n$ for $\lambda = 1.0$, $\lambda = 0.9$, $\lambda = 0.7$, $\lambda = 0.5$, and $\lambda = 0.0$.
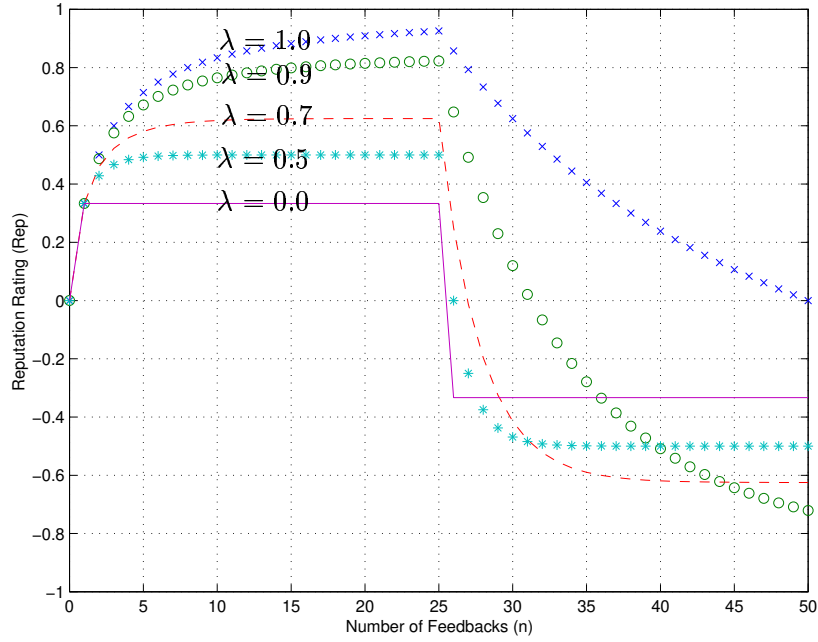


Figure 7: Reputation rating as a function of $n$ with varying feedback $v$ and forgetting factor $\lambda$.

It can be seen that the curves are identical to those on Figure 6 for $n \le 25$ where $v = 1$. For $n > 25$ where $v = -1$, the reputation rating drops quickly. Two phenomena can be observed when the forgetting factor is low (i.e. when feedback is quickly forgotten); Firstly the reputation rating reaches a stable value more quickly, and secondly the less extreme the stable reputation rating becomes.

# 4   Conclusion

We have presented the beta reputation system as a set of building blocks which makes it flexible and relatively simple to implement in practical applications. We have not tried to address the problem of immunity against agents changing identities other than assuming that some authentications mechanism is in place. We have also assumed a centralised approach, although it is possible to adapt the beta reputation system in order to become decentralised. In contrast to most other reputation systems that have been implemented or proposed, the beta reputation system has a sound theoretical basis in statistics, which together with its flexibility and simplicity makes it suitable for supporting electronic contracts and for building trust between players in e-commerce.

# References

[1] R. Axelrod. *The Evolution of Cooperation*. Basic Books, New York, 1984.

[2] George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Press, 1990.

[3] M. Ekstrom and H. Bjornsson. A rating system for AEC e-bidding that accounts for rater credibility. In *Proceedings of the ECCE ICT Symposium 2001*, Espoo, Finland, 2001.

[4] A. Jøsang. *Modelling Trust in Information Security*. PhD thesis, Norwegian University of Science and Technology, 1998.

[5] A. Jøsang. An Algebra for Assessing Trust in Certification Chains. In J. Kochmar, editor, *Proceedings of the Network and Distributed Systems Security Symposium (NDSS'99)*. The Internet Society, 1999.

[6] A. Jøsang. A Logic for Uncertain Probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–311, June 2001.

[7] L. Mui, P. Szolovits, and C. Ang. Collaborative sanctioning: applications in restaurant recommendations based on reputation (poster). In *Proceedings of the 5th International Conference on Autonomous Agents (AGENTS'01)*, pages 118–119, Montreal, Canada, 2001.

[8] B. Padovan, S. Sackmann, T. Eymann, and I. Pippow. A Prototype for an Agent based Secure Electronic Marketplace Including Reputation Tracking Mechanisms. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34)*, pages 2235–2244, Maui, Hawaii - Track 7, January 3-6 2001.

[9] M.A. Patton and A. Jøsang. Technologies for Trust in E-Commerce. In *Proceedings of the IFIP working conference on E-Commerce*, Salzburg, Austria, June 2001.

[10] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation systems. *Communications of the ACM*, 43(12):45–48, December 2000.

[11] J. Sabater and C. Sierra. REGRET: A reputation model for gregarious societies. In *Proceedings of the 4th Workshop on Deception, Fraud and Trust in Agent Societies, in the 5th International Conference on Autonomous Agents (AGENTS'01)*, pages 61–69, Montreal, Canada, 2001.

[12] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[13] E. Young. Not a pretty picture. *The Industry Standard (*available online), http://www.thestandard.com/article/0,1902,22875,00.html, March 26 2001.

[14] B. Yu and M.P Singh. A Social Mechanism of Reputation Management in Electronic Communities. In *Proceedings of the 4th International Workshop on Cooperative Information Agents*, pages 154–165, 2000.

[15] G. Zacharia, A. Moukas, and P. Maes. Collaborative Reputation Mechanisms in Electronic Marketplaces. In *Proceedings of the 32nd Hawaii International Conference on System Science*. IEEE, 1999.

[16] P.R. Zimmermann. *The Official PGP User's Guide*. MIT Press, 1995.