This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE P2P 2012 proceedings

Disconnection Punishment in Trust Bootstrapping: Benefits of Activity Stereotypes

Marc Sánchez-Artigas
Dept. of Computer Engineering and Maths, Universitat Rovira i Virgili, Spain
E-mail: marc.sanchez@urv.cat

***Abstract*—Trust-based systems have been proposed as means to fight against malicious agents in peer-to-peer networks. However, there still exist some issues that have been generally overlooked in the literature. One of them is the question of whether punishing disconnecting agents is effective. In this paper, we investigate this question for these initial cases where prior direct and reputational evidence is unavailable, what is referred in the literature as *trust bootstrapping*. First, we demonstrate that there is not a universally optimal penalty for disconnection and that the effectiveness of this punishment is markedly dependent on the uptime and downtime session lengths. Second, to minimize the effects of an inadequate selection of the disconnection penalty, we propose to incorporate predictions into the *trust bootstrapping process*. These predictions based on the current activity of the agents enhance the selection of potentially long-lived trustees, shortening the trust bootstrapping time when direct and reputational information is lacking.**

I. Introduction

Reputation and trust-based systems have been proposed for a number of applications, ranging from the selection of reliable servers in P2P networks to the detection of misbehaving nodes in mobile ad-hoc networks. But one of the main shortcomings that trust systems encounter in open systems is *how to interpret disconnection*. Disconnection affects quality of service (QoS). For example, in a P2P streaming service, QoS can be achieved as long as a continuous, uninterrupted data flow is maintained. For such a reason, streaming systems like ripple-stream [1] and reputation systems like [2], [3], [4] issue negative feedback for agents that are supposed to be providing the service but cannot do so now because they are logged off. The justification of this policy is that a peer can disconnect at any time and the trustor cannot ascertain whether the disconnection was *intentional* or not. For example, in a P2P file-sharing application, a peer may simply ignore queries despite owning the desired file, making the trustor believe that he is offline.

Punishing disconnecting agents, however, has its downsides. One that is not well understood is its connection with a greater risk of abuse. The more the importance given to disconnection, the less the *good* agents to request service from, because there are less with the sufficient availability to be potentially eligible for future interaction. Eventually, this can lead a trustor to *take a chance* on an unknown agent, or on an agent proven to be not completely trustworthy in the past, thereby increasing the risk of bad interaction.

The importance ascribed to disconnection is more critical in those situations where no prior direct and reputational evidence is available. This happens, for instance, when a new user enters the system for the first time, or when users form ad-hoc groups around a shared goal, which dissolve once that goal is reached. In these cases, the basic way of forming a confident opinion on users is through *direct interaction*. Since direct interaction with strangers maximizes the number of unsatisfactory experiences, the penalty imposed on disconnecting users plays an important role in *bootstrapping* trust when interacting with new agents.

For instance, consider the case that multiple unknown agents are offering the same service. Since a priori all agents have the same disposition to good action, a random trustee is chosen. If the trustee becomes unresponsive after completing a number of satisfactory transactions, the trustor will be confronted with the decision of whether to wait for its recovery or to take a chance on another agent. The latter is likely to occur if the penalty for disconnection is large. In that case, the trustor will maximize interaction but will be more exposed to abuse by the yet-to-be-known agents. On the contrary, if the disconnecting penalty is small, the trustee may return before getting low trustworthiness and continue to provide good service. This will minimize the risk of bad interaction but at the cost of an intermittent service.

The first contribution of this work is to analyze this tradeoff, and more generally, to quantify to which extent disconnecting penalty affects trust bootstrapping as a function of availability. To make the analysis mathematically tractable, we assume that T time units must elapse after disconnection in order to prefer an unknown agent. A smaller value for T represents a greater penalty, i.e., a higher probability for the trustor to take a chance on a new partner. Using this parameter, we develop a stochastic model to estimate the expected time to obtain the first *confident* trust evaluation on any of the strangers providing the service. This time we simply refer to as "*bootstrapping time*" is a good indicator of the efficacy of using a trust system. If this time is short, trustors will quickly form a useful impression to guide their interactions.

As a result of our analysis, we arrive at the conclusion that there is not a universally optimal penalty; the optimal penalty is too much dependent on the exact amount and type of churn. To address this issue, we propose to use peer activity as means to improve bootstrapping time while minimizing the effects of an inappropriate selection of the disconnecting penalty. To wit, a trustor may learn that the trustees downloading files between 1 to 4GB tend to have long sessions, and use this knowledge to select between two trustees based on their current downloading activity in the system. This concept is similar to the notion of stereotypes [5], [6], but applied to dynamics. Our results show that *activity* stereotypes are of great help in bootstrapping trust in the problematic initial cases discussed here.

II. Analysis of Punishing Disconnection

Model. To turn this into a generic analysis, we simply assume that the trustor must interact at least ℓ times with any trustee to collect enough *direct* evidence to feel confident in the resulting *local trust value*. That is, the value of ℓ marks the point where uncertainty about the result of the next interaction (positive or negative) is low. While the exact value of ℓ will vary from one system to another, note that our approach will remain valid for

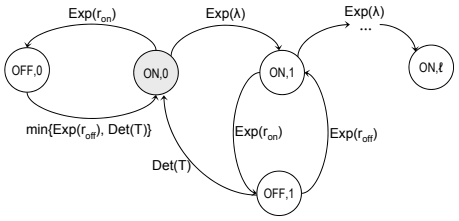


Fig. 1. State diagram for the semi-Markov process $\{Y(t), t \geq 0\}$.

many trust and reputation systems. Based on this parameter, we define the *trust bootstrapping time* as:

Definition 1. *Given a group of unknown agents $\mathcal{U}$ providing a service, we define the trust bootstrapping time τ_ℓ as the time to complete the first ℓ transactions with any of the agents in $\mathcal{U}$.*

For analytical tractability, we assume that transactions occur *immediately one after another*, according to a Poisson process with rate λ. As a result, transactions exhibit a similar duration, making unnecessary to calculate the gain in trust in proportion to the amount of work done.

For simplicity, the alternating ON/OFF behavior of trustees is modeled with the help of a 2-state continuous-time Markov chain (CTMC) with transition rates r_{on} and r_{off} in the ON and OFF states, respectively. In systems dominated by user-driven interruptions such as Maze or Kad, it has been recently verified that this simple CTMC provides a good approximation to user behavior [7].

To model disconnection punishment, we assume that T time units must elapse after disconnection for the current trustee to get lower trustworthiness than a stranger, i.e., the default trust value assigned to an unknown agent. Smaller values of T mean a greater penalty, i.e., a higher probability for the trustor to take a chance on a new agent. Larger values of T, on the contrary, reduce the risk of bad response. If the current trustee is giving good service, a larger T trades off longer interruptions in the service against the risk of switching to an unknown agent, who can be malicious. Consequently, by varying the value of T we can analyze the tradeoff between risk and QoS.

For tractability, we assume that the result of each transaction is positive to initiate a new one with the current trustee. Notice that if trustees behave badly, the number of agent switches will be greater due to negative responses. As a result, the expected bootstrapping time will be longer. In practice, this assumption makes our results conservative but accurate enough to measure the impact of disconnections. In fact, our analytical results are in good agreement with our simulations reported in Section IV. With this assumption, the trustor switches to a new agent only when T runs out, which simplifies the stochastic chain.

We have used the model of decision most commonly found in the literature that involves selecting the *most trusted agent*.

The state transition diagram is given in Fig. 1. States (ON, i) and (OFF, i) represent the case where the number of completed transactions is $i \geq 0$ and the current trustee is ON and OFF, respectively. In state (ON, i), the process can jump into either state (ON, $i+1$), which represents that a new transaction has ended, or state (OFF, i), which implies that the trustee is now offline. In this state, the process can jump into state (ON, 0) if disconnection time exceeds T, which implies a *trustee switch*, or to state (ON, i) otherwise. The state of the process at time 0 is of course (ON, 0).

The kernel $\mathbf{Q}(t) = \left[Q_{v,j}^{\xi,i}(t)\right]$ of the process, say $\{Y(t)\}_{t\geq 0}$, is as follows:

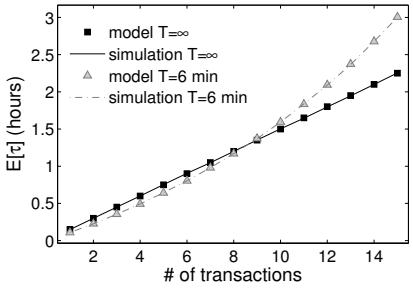


Fig. 2. $\mathbb{E}[\tau_\ell]$ plotted against ℓ. Mean ON time $\mathbb{E}[L] = 1$ hour, mean OFF time $\mathbb{E}[D] = 0.5$ hours, with $\lambda = 10$ transactions/hour.

$$
\begin{aligned}
Q_{\text{ON},0}^{\text{OFF},0}(t) &= \Pr\{\textit{trustee recovers before or at time } t,\ t < T,\\
&\quad \textit{with no transactions completed}\}\\
&= 1 - e^{-r_{off}t} + e^{-r_{off}t}u(t-T).\\
Q_{\text{OFF},i}^{\text{ON},i}(t) &= \Pr\{\textit{trustee logs off before transaction } i+1\\
&\quad \textit{finishes}\}\\
&= \tfrac{r_{on}}{r_{on}+\lambda}\left(1 - e^{-(r_{on}+\lambda)t}\right), \quad \forall i = 0, ..., \ell-1.\\
Q_{\text{ON},i+1}^{\text{ON},i}(t) &= \Pr\{\textit{transaction } i+1 \textit{ completed before the}\\
&\quad \textit{trustee goes to OFF state}\}\\
&= \tfrac{\lambda}{r_{on}+\lambda}\left(1 - e^{-(r_{on}+\lambda)t}\right), \quad \forall i = 0, ..., \ell-1.\\
Q_{\text{ON},i}^{\text{OFF},i}(t) &= \Pr\{\textit{trustee goes to ON state before}\\
&\quad \textit{or at time } t \textit{ and } t < T\}\\
&= 1 - e^{-r_{off}t} - \left(e^{-r_{off}L} - e^{-r_{off}t}\right)u(t-T),\\
&\quad \forall i = 1, ..., \ell-1.\\
Q_{\text{ON},0}^{\text{OFF},i}(t) &= \Pr\{\textit{trustee does not go to ON state}\\
&\quad \textit{at time } t \textit{ and } t \geq T\}\\
&= e^{-r_{off}T}u(t-T), \quad \forall i = 1, ..., \ell-1.
\end{aligned}
$$

where $u(t-T)$ is the unit step function at T.

Because we are interested in the time to complete the first ℓ transactions with any of the unknown trustees, it can be easily verified that the trust bootstrapping time τ_ℓ corresponds to the first-hitting time of process $\{Y(t)\}_{t\geq 0}$ onto state (ON, ℓ) given that $Y(0) = (\text{ON}, 0)$:

$$\tau_\ell = \inf\{u > 0 : Y(u) = (\text{ON}, \ell) | Y(0) = (\text{ON}, 0)\}.$$

Following the derivation explained in [8], we can find the average bootstrapping time $\mathbb{E}[\tau_\ell]$ and determine the influence of *disconnection punishment* on trust evaluation when no prior evidence can be found:

Theorem 1. For user ontimes L with CDF $1 - e^{-r_{on}x}$, user offtimes D with CDF $1 - e^{-r_{off}x}$ and threshold T, the mean time to complete the first ℓ transactions with a trustee and feel confident in the resulting trust value is given by:

$$
\begin{aligned}
\mathbb{E}[\tau_\ell] &= \frac{1}{r_{off}\lambda^\ell}\sum_{i=0}^{\ell}\left(r_{on}e^{-Tr_{off}}\right)^{\ell-i}S_{i,\ell}, \qquad (1)\\
S_{i,\ell} &= \left(\eta_{\ell,i}r_{off}\lambda^{i-1} - \epsilon_{\ell,i}\lambda^i + \epsilon_{\ell,i}\lambda^i e^{Tr_{off}}\right),
\end{aligned}
$$

where $r_{on} = 1/\mathbb{E}[L]$, $r_{off} = 1/\mathbb{E}[D]$, and $\eta_{\ell,i}$ and $\epsilon_{\ell,i}$ satisfy the recurrence relations $\eta_{\ell,i} = \eta_{\ell-1,i} + \eta_{\ell-1,i-1}$ for all $i > 1$, $\epsilon_{\ell,i} = \eta_{\ell,i+1}$ for all $i < \ell$. The initial conditions are $\eta_{\ell,0} = 0$, $\eta_{\ell,1} = 1$, $\eta_{\ell,\ell} = \ell$ and $\epsilon_{\ell,\ell} = 0$.

Results. The first observation to be made is that the influence of churn can be predicted by taking Eq. (1) to the limit ($T \to \infty$), which yields equation:

$$\mathbb{E}[\tau_\ell] = \frac{\ell}{\lambda}\left(\frac{r_{off} + r_{on}}{r_{off}}\right). \quad (2)$$

This equation has a straightforward interpretation: the average bootstrapping time is *inversely proportional* to the steady-state user availability $\mathcal{A} = \frac{\mathbb{E}[L]}{\mathbb{E}[L]+\mathbb{E}[D]} = \frac{r_{off}}{r_{on}+r_{off}}$. Since there is no punishment in this case ($T \to \infty$), Eq. (2) points out that the time spent in accruing enough supporting evidence to make a confident trust evaluation depends basically on the odds for the initial random trustee to stay connected. Hence, for a cautious trustor with a larger rather than shorter T, the minimization of the risk when interacting with strangers causes trust evaluation *to be highly dependable on agent availabilities*. Consequently, if availabilities are low, our analysis prompts us to suggest that it is advantageous *to choose another agent after a short period of inactivity*. For instance, if agent availability is of $\mathcal{A} = 0.5$, Eq. (2) tells us that the bootstrapping time $\mathbb{E}[\tau_\ell]$ is doubled in the absence of disconnection punishment.

However, the key question is whether an appropriate amount of disconnection punishment can do it better. As just discussed above, it appears at first glance that an *aggressive* punishment should decrease the trust bootstrapping time and favor a more continuous service (less inactivity periods). While the latter is, in general, correct, *the former is not necessarily true*. Contrary to intuition, a short T may increase the trust bootstrapping time $\mathbb{E}[\tau_\ell]$ if, as commonly happens, the extended design principle that good behavior should increase trust slowly is applied. For instance, in PET [2], the penalty for "*No Response*" is 3 times that of good action.

An example of this is shown in Fig. 2. In the figure, $\mathbb{E}[\tau_\ell]$ is plotted against ℓ (the estimate from Eq. (1) is compared to simulations). In this case, the mean ON time $\mathbb{E}[L] = 1$ hour, the mean OFF time $\mathbb{E}[D] = 0.5$ hours, with a transaction rate λ of 10 transactions per hour. As shown in the figure, when ℓ increases, the time required for an aggressive punishment ($T =$ 6 min.) to form a confident opinion becomes greater than when imposing no penalty on disconnecting trustees ($T = \infty$). The main reason is that departed agents present a strong tendency to return sooner than later [9]; therefore, an aggressive strategy to maximize interaction might penalize in excess accuracy in trust evaluation. This suggests that T should be ideally chosen based on the distribution of downtime, defined by the interval between the moment an agent disconnects and its next arrival, which is hard to achieve in practice.

Overall, our analysis prompts us to conclude that *there does not exist a universally "optimal" penalty for disconnection* and that *the effectiveness of this punishment is strongly dependent on the downtime session lengths*.

III. Activity Stereotypes

The lack of a global optimal disconnecting penalty demands new techniques to improve trust bootstrapping while protecting the system from the churn of the P2P network. One way to do so is to incorporate predictions on peer uptimes into the trust bootstrapping process. Among the possible solutions, we focus on the current activity of an agent as a predictive mechanism. For instance, in a file-sharing application, a trustor may learn that the peers downloading files between 1GB to 4GB tend to have long sessions, and use this knowledge to choose between two agents based on their downloading activity in the system. By ascribing peer selection to learned classes of agent activity, a trustor could employ prior experiences in similar activities to protect trust evaluation from the churn of the P2P system. This concept is similar to the notion of *stereotypes* firstly proposed in [6] and later in [5], but applied to ON-OFF dynamics.

Since our focus is on demonstrating the potential of making use of activity stereotypes to avoid the need to make a random partner selection, we do not involve ourselves on issues such as how this information is obtained or how the current activity of each partner is monitored. These types of issues are left for future work. We simply assume that the availability history for the peers realizing a given activity A_i is maintained somehow, and this knowledge can be used to predict the uptime duration of a partner performing a similar activity.

An important requirement of activity stereotypes is that they are meant to complement, not replace, direct evidence about an individual when it is available. While activity stereotypes may facilitate useful predictions in initial conditions, they are based on empirical generalizations, and should carry less weight than direct observation. Similar to [5], we fulfill this requirement by adapting the *default trust* in unknown agents to the behavior of the majority of agents performing a given activity.

More formally, the objective of our mechanism is to identify a function f that maps the activity vector of an agent $\vec{A}$ to an estimate on service continuity S which increases or decreases the default trust in that agent. This representation enable us to assess the potential of activity stereotypes for any general trust evaluation mechanism.

A. Trust Model

Regardless of the underlying model, the key requirement of our approach is that the estimates that function f produces are compatible with the trust model being used. For this reason, we describe here the concrete trust model we use to demonstrate the potential of activity stereotypes on bootstrapping trust.

Specifically, we adopt the model proposed in [10] and based on Subjective Logic. The reason for using this model is that by mapping activity stereotypes to base rates, their effect reduces as more direct evidence is observed. In addition, it provides an intuitive way for the trustor *to measure the quantity of evidence that supports belief towards a given agent*, what is known as *uncertainty*.

Representation. In this model, an *opinion* held by a trustor x about an agent y is represented in Subjective Logic as a tuple $w_y^x = \langle b_y^x, d_y^x, u_y^x, a_y^x \rangle$, where values b_y^x, d_y^x, u_y^x, a_y^x express the degree of *belief*, *disbelief*, *uncertainty* and *base rate* (or a priori degree of belief), respectively. These values satisfy the relation $b_y^x + d_y^x + u_y^x = 1$ with $b_y^x, d_y^x, u_y^x, a_y^x \in [0, 1]$. Belief expresses to which extent x believes that interaction with y will result in a positive outcome. Uncertainty u_y^x is caused by the absence of evidence to support either belief or disbelief, so that an opinion based on 100 transactions has a greater certainty than another based on just 1 observation.

Evidence Aggregation. A trustor bases his opinions on evidence, which is obtained by interacting with other agents,

and can be positive or negative. A body of evidence held by a trustor x is a pair $\langle r_y^x, s_y^x \rangle$, where r_y^x is the number of positive transactions received from y, and s_y^x is the number of negative experiences. Using these two parameters, an opinion is produced as follows:

$$\begin{aligned} b_y^x &= r_y^x/\big(r_y^x + s_y^x + 2\big); \quad d_y^x = s_y^x/\big(r_y^x + s_y^x + 2\big) \\ u_y^x &= 2/\big(r_y^x + s_y^x + 2\big). \end{aligned} \tag{3}$$

Observe that Eq. (3) guarantees that uncertainty decreases as more evidence is collected. Alternatively, evidence could be obtained from third parties who had interacted with a specific individual before. However, since we examine the problem of trust establishment when no historical information is available, evidence is acquired *first hand* by each trustor.

To treat '*No Response*' as a bad action like in PET [2] and [3], so that the peers joining and leaving the system frequently get low trustworthiness, we classify negative experiences into '*Bad Behavior*' and '*No Response*', and calculate s_y^x as linear combination of the observed frequencies of each type:

$$s_y^x = \gamma_{\mathtt{B}} \cdot s_{y:\mathtt{B}}^x + \gamma_{\mathtt{N}} \cdot s_{y:\mathtt{N}}^x, \tag{4}$$

where $s_{y:\mathtt{B}}^x$ is the number of wrong or malicious transactions, $s_{y:\mathtt{N}}^x$ is the number of transactions that got no response, and $\gamma_{\mathtt{B}}$ and $\gamma_{\mathtt{N}}$ are the weights attached to each type of negative action. These weights are used to assign different levels of importance to each type of negative experience. In our simulations, we will set $\gamma_{\mathtt{B}} = 1$ and will vary the value of $\gamma_{\mathtt{N}}$ to measure the impact of disconnection punishment on trust bootstrapping times.

Trust Metric. In this model, a single-valued trust metric, useful for ranking agents, can be derived from a particular opinion w_y^x as follows:

$$P\left(w_y^x\right) = b_y^x + a_y^x \cdot u_y^x, \tag{5}$$

where the resultant trust value corresponds to the probability expectation value $P\left(w_y^x\right)$ for w_y^x. The base rate a_y^x represents the a priori degree of trust x has about y before any evidence is received. It determines the effect that the parameter u_y^x will have on the resultant trust value. The default value of a_y^x, and hence *default trust*, is 0.5, which signals that in the absence of evidence, both positive and negative outcomes are considered as equally likely to occur. In this case, $P\left(w_y^x\right) = 0.5$, which is the least informative value about an agent. Values of $a_y^x > 0.5$ will result in more uncertainty being converted to belief, and conversely disbelief for $a_y^x < 0.5$.

Reputation. Reputation in probabilistic trust systems is usually calculated by aggregating evidence from trustful providers [2], [10]. Since we study the effects of disconnection in those initial situations whereby previous direct and reputational evidence is unavailable, the result of the aggregation of evidence provided by the mutually unknown agents will lead to weak reputations, as some of them may be unreliable or malicious. As a result, for informed peer selection we make only use of the local trust values computed at each trustor.

B. Stereotype Function

To incorporate our predictions into the trust bootstrapping process, we use the base rate. That is, for a given trustee y, the base rate $a_y^x = f(\vec{A}_y)$. When no evidence has been accrued for trustee y we have maximum ambiguity, i.e. $w_y^x = \langle 0, 0, 1, 0.5 \rangle$. In this case, it is easy to see that a_y^x alone determines the value of $P\left(w_y^x\right)$. However, as more evidence is obtained, the value of u_y^x diminishes, and therefore the weight carried by a_y^x in the trust value also decreases. This fulfills the requirement that the effect of our initial predictions must decay as direct evidence is accrued. We refer to this condition as Requirement 1.

Another key observation to be made about our approach is that activity stereotypes are useful to form a tentative estimate of the '*No Response*' component of trust, not of the whole trust evaluation. This means that the increase of the base rate above *default trust* must never prevent the estimated trust value from reducing rapidly if the trustee starts to misbehave. Otherwise, a malicious agent can perform an activity where individuals tend to stay longer to attract trustors, and then start to behave badly. Although this requirement is subsumed by Requirement 1, we determine an upper bound on the base rate that guarantees that predicted trust values drop below default trust after the failure of at most I consecutive transactions, where I is an expression of the risk of prediction. If I is small, a malicious trustee will be rapidly detected and stereotypical information will be less prone to manipulation. For brevity, we refer to this requirement as Requirement 2. Now let a_{def} be the default or neutral trust value (usually 0.5). Then,

Lemma 1. Given default trust $a_{def} \in [0, 1]$, and risk tolerance I, $I \in \mathbb{N}_1$, the satisfaction of Requirement 2 requires the base rate a_y^x to be:

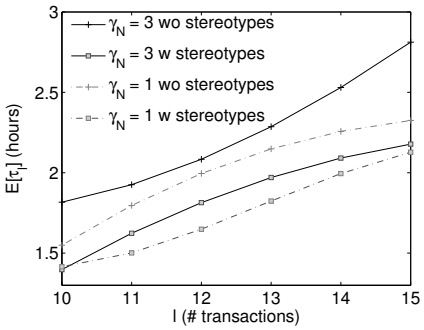
$$a_y^x \leq \min\left(a_{def}\frac{(I+2)}{2}, 1\right). \tag{6}$$

From Eq. (6), it is easy to see that $I = 2$ consecutive bad interactions are enough to drop the predicted trust value below $a_{def} = 0.5$ even though $a_y^x = 1$, which shows that estimates on the '*No Response*' component of trust do not condition the whole trust formation process in initial cases.

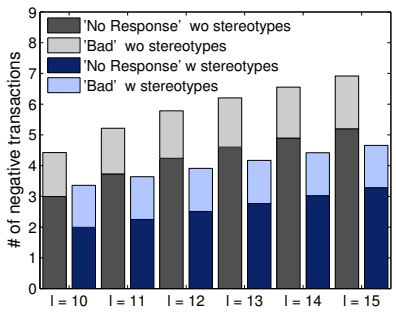
Based on the above observations, we are ready to discuss on the shape of the stereotypical function f. This function takes as input a vector of activities $\vec{A}_y$ being currently carried out by an unknown agent y and returns a prediction of its susceptibility to participate continually.

To compute such an estimate, for each activity A_i in $\vec{A}_y$, the trustor first obtains the probability that an agent of activity A_i stays connected for a duration of ℓ transactions, i.e., ℓ/λ. We refer to this probability simply as $p_i^{\ell/\lambda}$. We assume that this value is obtained by asking one of the trusted parties who keep a record of the uptime session durations of the prior agents that performed such an activity. Note that from the uptime session durations of agents, probability $p_i^{\ell/\lambda}$ can be easily obtained by computing $F_{A_i}^c\left(\frac{\ell}{\lambda}\right) = \Pr\left\{L > \frac{\ell}{\lambda} | A_i\right\}$, where $F_{A_i}^c(\cdot)$ denotes the empirical complementary uptime distribution observed in the agents that carried out activity A_i. For the acquisition of uptime session durations, we assume the existence of a secure monitoring protocol for peer availability such as AVMON [11] tailored to classify by type of activity.

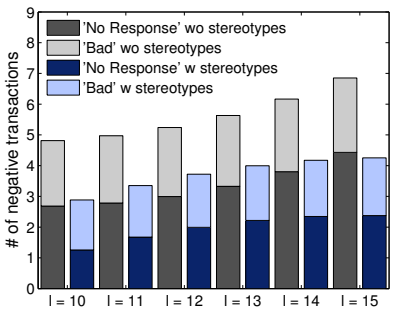
From probabilities $p_i^{\ell/\lambda}$, the trustor picks the maximum of them and normalizes it to the range $[a_{def}, a_{max}]$, where a_{max} is the threshold on the base rate calculated from Eq. (6). The reason of the normalization is to avoid favoring the new agents for which no activity is known, as these agents get assigned the default trust value. In this way, agents from which prior knowledge on activities can be leveraged will always be preferred over trustees for which no activity is known, thereby encouraging agents to participate in the system.



(a) Mean bootstrapping time as ℓ is varied.



(b) Number of negative interactions for $\gamma_{\mathtt{N}} = 1$.



(c) Number of negative interactions for $\gamma_{\mathtt{N}} = 3$.

Fig. 3. Effectiveness of activity stereotypes when used together with a relatively straightforward probabilistic trust model.

Putting all pieces together, stereotyping function f is given by:

$$f(\vec{A}_y) = \max\left\{p_i^{\ell/\lambda}\right\}_{A_i \in \vec{A}_y} (a_{max} - a_{def}) + a_{def}. \quad (7)$$

IV. Evaluation

In evaluating our proposal, we simulated a system composed of $N_{groups} = 250$ groups in which there is a trustor who wants to receive service from 10 agents from which no trust evidence is available. The goal of the trustor in each group is to interact with 10 new agents while minimizing risk. For each group, we record the time needed for the trustor to form the first reliable opinion in an agent in the group, a.k.a., the trust bootstrapping time. In addition, we count the number of negative interactions received up to this time, including the transactions that got no response.

Each of the $2,500$ trustees is assigned an activity profile that specifies how long it will be connected and disconnected. This activity profile can be composed of up to two activities. Prior agents realizing the first activity A_1 exhibited uptime durations drawn from $F_{A_1}(x) = 1 - e^{-0.2x}$, and offtime durations drawn from $G_{A_1}(x) = 1 - e^{-x}$, which corresponds to a mean uptime and offtime of 5 and 1 hours, respectively. On the contrary, the agents that performed activity A_2 in the past presented a high turnover rate with an average uptime and offtime of 15 and 30 minutes, respectively. In our simulations, only 500 out of the $2,500$ trustees are assigned activity A_1 to make it difficult for non-stereotype peer selection to discover long-lived trustees. In addition, 50% of the trustees are specified to act maliciously. This means that 250 of the long-lived trustees will misbehave, thereby requiring our approach to react against the misbehavior of long-lived trustees (Requirement 2). The rest of parameters are $\lambda = 10$ interactions/hour, $a_{def} = 0.5$ and $a_{max} = 0.75$.

The results are illustrated in Fig. 3 for two distinct values of disconnection penalty: $\gamma_{\mathtt{N}} = 1$ (small punishment) and $\gamma_{\mathtt{N}} = 3$ (aggressive punishment). Besides the expected conclusion that informed peer selection using activity stereotypes performs the best in all cases, two major observations should be made about the results. The first is that, contrary to intuition but consistent with our analysis in Section II, an aggressive punishment can increase the trust bootstrapping time, instead of reducing it, as depicted in Fig. 3(a). This result arises because in our scenario the trustees present a marked tendency to return sooner, which makes it preferable for a trustor to wait for the trustee to come back rather than to take a chance on a new agent. In this sense, *activity stereotypes help to reduce the influence of an improper selection of the disconnection penalty*.

The other key observation is the effect that the disconnection penalty has on the tradeoff between risk and service continuity (QoS). As shown in Fig. 3(b) and 3(c), aggressive punishment of disconnection, although it reduces the occurrence of service intermittences, increases the risk of participating in a malicious transaction, and vice versa for low penalty. In this regard, *the use of stereotypes can only be useful to decrease the number of service interruptions, not the risk of interaction which depends on the magnitude of the penalty*.

V. Conclusions

In this work, we have investigated how affects disconnection penalty the process of trust formation for these initial situations where prior evidence is unavailable. First, we have analytically proven the lack of a universally optimal penalty and shown its dependence on the disconnection pattern of agents. Finally, we have introduced a mechanism that leverages prior knowledge on peers' activities to enhance the trust bootstrapping process, making it less dependent on the way disconnection is treated.

Acknowledgements

This work has been partly funded by the Spanish Ministry of Science and Innovation through projects DELFIN (TIN-2010-20140-C03-03) and RealCloud (IPT-2011-1232-430000).

References

[1] W. Wang et al., "Ripple-stream: Safeguarding p2p streaming against dos attacks," in *Proc. IEEE ICME'06*, 2006, pp. 1417–1420.
[2] Z. Liang and W. Shi, "PET: A PErsonalized Trust Model with Reputation and Risk Evaluation for P2P Resource Sharing," in *Proc. HICSS'05*, 2005, pp. 201b–201b.
[3] N. Fedotova and L. Veltri, "Reputation management algorithms for dht-based peer-to-peer environment," *Computer Communications*, vol. 32, no. 12, pp. 1400–1409, 2009.
[4] X. Li et al., "A multi-dimensional trust evaluation model for large-scale p2p computing," *J. Parallel Distrib. Compu.*, vol. 71, no. 6, pp. 837–847, 2011.
[5] C. Burnett, T. J. Norman, and K. Sycara, "Bootstrapping trust evaluations through stereotypes," in *Proc. AAMAS'10*, 2010, pp. 241–248.
[6] X. Liu, A. Datta, K. Rzadca, and E.-P. Lim, "Stereotrust: a group based personalized trust model," in *Proc. CIKM'09*, 2009, pp. 7–16.
[7] Z. Yang et al., "Exploring peer heterogeneity: Towards understanding and application," in *Proc. IEEE P2P'11*, 2011, pp. 20–29.
[8] V. G. Kulkarni, *Modeling and analysis of stochastic systems*. London, UK, UK: Chapman & Hall, Ltd., 1995.
[9] D. Stutzbach and R. Rejaie, "Understanding churn in peer-to-peer networks," in *Proc. ACM IMC'06*, 2006, pp. 189–202.
[10] A. Jøsang, R. Hayward, and S. Pope, "Trust network analysis with subjective logic," in *Proc. ACSC'06*, 2006, pp. 85–94.
[11] R. Morales and I. Gupta, "Avmon: Optimal and scalable discovery of consistent availability monitoring overlays for distributed systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 20, no. 4, pp. 446–459, 2009.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE P2P 2012 proceedings

Summary Review Documentation for

"Disconnection Punishment in Trust Bootstrapping: Benefits of Activity Stereotypes"

Authors: Marc Sánchez-Artigas

Summary Review

The paper deals with an interesting topic that bootstraps the trust with P2P peers. Disconnection punishment is one mechanism that could screen selfish/malicious peers and that could identify a set of trustworthy peers, but the mathematical model in the paper finds that the universally optimal penalty is hard to be obtained since it is too dependent on the amount and the type of churn. Instead, it adjusts the peer activity stereotype to incorporate peer uptime to trust bootstrapping.

I like the rigorous treatment of disconnection punishment, and its findings are quite useful and interesting. However, it is hard to follow the material since the paper does not define terms or clarify the concepts clearly. For example, some of the terms in Section III are vague to me. How do you define a positive/negative opinion? What are positive/negative transactions and a positive outcome? In equation (3), what is the meaning of the number, 2? What are wrong or malicious transactions? It can be simply me that do not understand the standard terms in this area, but it would be beneficial to define these terms to make the paper more approachable.

Strengths: The strengths of the paper include (i) the paper is succinct and fairly easy to follow; (ii) it exploits non-historical/reputation information to determine whether to "trust" a node for an interaction by applying stereotypes, (iii), it includes a rigorous analysis on the effect of disconnection punishment, (iv) the idea of bootstrapping trust based on response rates seems sensible.

Weaknesses: The weaknesses pointed out by the review process include: (i) the scope of the paper is narrow (although to be fair this is a short paper), (ii) stereotype based trust has been used in the context of P2P back-up storage systems (see detailed comments below). While I do think that the way it is used in this paper is somewhat different from the other work, and hence has adequate novelty to merit for a short paper, the authors do seem unaware of such closely related works (which they need to comment/distinguish themselves with), (iii) the paper does not analyze any real systems or data derived from real systems, (iv) the analysis itself does not seem particularly novel, and and (v) some of the assumptions in the formal analysis need proper justification.

Response from the Authors

All the reviewers agree that the strength of the paper is the study of an important aspect of trust in P2P systems, i.e, the punishment for disconnections, followed by a promising solution to be further explored.

The paper is well balanced in this regard, but raised several questions especially with respect to the mathematical analysis of the effect of disconnection punishment. While one reviewer was genuinely enthusiastic and only suggested the inclusion of a missing reference, two reviewers were particularly concerned with the assumptions adopted in the analytical model as well as with some technical definitions. All these concerns have been addressed by clarifying the definitions and providing better justifications for the assumptions. For instance, the use of a Poisson process to model the occurrence of interactions is now justified by saying that a Poisson process was adopted to ensure that transactions exhibit a similar duration, making unnecessary to calculate the gain in trust in proportion to the amount of work done. Similarly, the assumption that the result of each transaction is sufficiently positive to initiate a new transaction with the current trustee, which was adopted for mere tractability, it is now motivated by the drastic simplification of the stochastic process it entails while not representing a significant loss of accuracy in the estimated impact of disconnections. As argued in the paper, accuracy is not lost because receiving bad transactions from trustees causes more agent switches, thereby increasing the trust bootstrapping time.

Further, the notion and the implications of the trust bootstrapping time have been clarified. Especially, it has been signaled that a short bootstrapping time is indeed necessary to allow trusters to quickly form a useful impression on one another in order to guide future interactions.

Finally, typos and weird sentences have been corrected and re-worded as pinpointed by the reviewers.