A Game Theoretic Model for Digital Identity and Trust in Online Communities

Tansu Alpcan
Deutsche Telekom Labs
Technical University Berlin
Berlin, Germany
alpcan@sec.t-labs.tu-berlin.de

Cengiz Örencik
Sabanci University
Istanbul, Turkey
cengizo@su.sabanciuniv.edu

Albert Levi, Erkay Savaş
Sabanci University
Istanbul, Turkey
-
levi@sabanciuniv.edu
erkays@sabanciuniv.edu

ABSTRACT

Digital identity and trust management mechanisms play an important role on the Internet. They help users make decisions on trustworthiness of digital identities in online communities or e-commerce environments, which have significant security consequences. This work aims to contribute to construction of an analytical foundation for digital identity and trust by adopting a quantitative approach. A game theoretic model is developed to quantify community effects and other factors in trust decisions. The model captures factors such as peer pressure and influence of community leaders. The existence and uniqueness of a Nash equilibrium solution is studied and shown for the trust game defined. In addition, synchronous and asynchronous update algorithms are shown to converge to the Nash equilibrium solution. A numerical analysis is provided for a number of scenarios that illustrate the interplay between user behavior and community effects.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Communications Applications; G.1.6 [**Optimization**]: Gradient methods; K.4.4 [**Electronic Commerce**]: Security

Keywords

Digital identity, game theory, trust, online communities

1. INTRODUCTION

Digital identity constitutes one of the building blocks of the World Wide Web for all types of activities ranging from social networking to e-commerce. During the explosive growth phase of the Web, a variety of digital identity and trust management mechanisms have been developed organically to satisfy the emerging needs. However, most of these existing solutions have been either ad-hoc or heuristic in nature [1]. An analytical foundation for digital identity and trust can play an important role in continuing growth of interactive nature of Web services and social networks.

Game theory provides a rich set of mathematical abstractions and frameworks suitable for a quantitative treatment of digital identity and trust problems. Since it studies multi-person decision making with conflicting interests, game theory naturally supports development of an analytical foundation in this area. Quantitative models are useful for generalization of problems, combining the existing ad-hoc schemes, and opening doors to novel solutions. Hence, they bring a unique advantage over heuristic schemes which are problem specific and error prone. This paper presents such a quantitative model that formalizes community interactions in the context of trust in online environments. The objective is to gain additional insights to basic principles and develop algorithms that address existing and future digital identity and trust-related problems.

Digital trust and reputation are two concepts that are closely related to each other. An individual often decides to trust a digital identity or not based on the reputation of that identity. Therefore, reputation of a digital identity can be seen as a aggregate metric which is a function of the trust of community members in that digital identity. Online environments allow for quick dissemination and sharing of such trust decisions (user opinions) through rating systems. It is worth to noting that the term "trust" is used in this paper in a social context, in the sense of trusting a digital identity. This should be distinguished from trust in "trusted computing" or "trusted systems", where the term denotes consistent behavior enforced by hardware in the former and reliance upon a system to enforce a specified security policy in the latter.

The game theoretic model in this paper differentiates from earlier studies [2–7] by taking into account community influences and interactions explicitly. Factors such as peer pressure, personality traits such as timidness or reluctance to pass judgment, and influence of community leaders are investigated in a noncooperative game setting. The players (users) take part in a digital trust management system where they explicitly share their opinions on an external digital identity (e.g. seller in e-commerce). After a dynamic evaluation process, the resulting opinion is a mixture of their own individual assessment and community influences. The effect of various parameters on the final outcome as well as equilibrium and convergence properties of the iterative process are rigorously studied. The approach and results are illustrated and discussed based on three example scenarios.

The main contributions of this paper include: (a) a novel game theoretic model of community effects on trust in digital identities that captures factors such as peer pressure and influence of community leaders (b) rigorous study and proof of existence and uniqueness of a Nash equilibrium in the noncooperative digital trust game (c) global convergence analysis of parallel update algorithms for solving the trust game in a distributed manner.

ASIACCS'10 April 13–16, 2010, Beijing, China.
Copyright 2010 ACM 978-1-60558-936-7 ...$10.00.

2. DIGITAL TRUST GAME

Consider a set of **agents**, $\mathcal{A} := \{a_1, \ldots, a_i, \ldots, a_N\}$, which can represent users of a social network (e.g. *Facebook* or *Slashdot*) or participants in an e-commerce environment such as the one provided by *Amazon* or *Ebay*. For simplicity, each agent is associated with a single digital identity which is issued by a digital identity provider. This role is customarily played by the respective owner of the social networking or e-commerce site itself as in the case of *Amazon, Facebook,* or *Ebay*.

The **digital trust game** is played among N agents in the set $\mathcal{A}$, who evaluate a single given identity or seller s over a certain finite time interval. In the remainder of the paper the terms agent, user, and buyer as well as the terms evaluated identity and seller will be used interchangeably without any loss of generality. It is assumed here that the seller has a stationary *initial reputation* over this time window. The perceived *initial image* of the seller by individual agents may, however, vary according to personal experiences and observations. The digital trust game allows agents to form new *opinions* on the seller by sharing their evaluations and may result in a *community reputation* (aggregate trust) that differs from the initial reputation.

Given the initial reputation of the seller, $r_s \in \mathbb{R}$, the initial image (or trust level), $e_i \in \mathbb{R}$ perceived by an agent a_i can be considered as a noisy measurement of r_s and defined by

$$e_i := r_s + n_i. \tag{1}$$

The *bias term*, n_i, captures the individual variation in initial opinion of agent i on the seller. This may be a result of varying personal experiences or observational limitations and distortions. Depending on the specific system, the vector $n = [n_1, \ldots, n_N]$ can be modeled as additive (zero-mean) Gaussian noise.

Using the initial image e_i as a starting point, an agent a_i forms an **opinion (trust)**, $x_i \in \mathbb{R}$, of the seller after exchanging information with the rest of the community. The individual opinion or trust, x_i, is influenced by various community effects as well as individual properties of the agent. The opinions of all the agents represented by the vector

$$\mathbf{x} = [x_1, \ldots, x_N] \in \mathcal{X} \subset \mathbb{R}^N$$

define the decision space of the digital trust game. In many cases, the opinions are time-dependent as they are formed over time through an iterative update process.

In the game, $x_i = 0$ corresponds to a neutral or default opinion of agent a_i on the seller. Consequently, the positive values, $x_i > 0$ represent a positive opinion and negative ones, $x_i < 0$, a negative opinion. The same convention is also applied to the variables r_s and e, which have similar interpretations.

The agents' opinions are not only a function of the initial reputation and image but also of factors capturing community influences. The decision process of an agent a_i can be modeled by the minimization of a well-defined cost function that quantifies the factors affecting the opinion of the agent. One possible cost function of agent a_i adopted in this paper is

$$J_i(x_i, \mathbf{x}_{-i}) := \frac{\alpha_i}{2}x_i^2 + \frac{\beta_i}{2}\left(x_i - \frac{1}{N-1}\sum_{j \neq i} x_j\right)^2 + \frac{\gamma_i}{2}(x_i - e_i)^2, \tag{2}$$

where $0 \leq \alpha_i, \beta_i, \gamma_i \leq 1$, $\alpha_i + \beta_i + \gamma_i = 1 \quad \forall i$, and $x_{-i} := [x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N]$. It is naturally possible to consider different types of cost functions. This particular one is chosen for its nice analytical properties as a first order approximation.

The **first term**, $\alpha_i x_i^2$, in the cost function (2) quantifies the timidness of agent a_i. The term quadratically penalizes any positive or negative opinion of the agent forcing it to the neutral or zero opinion. Agents with different properties can be represented by choosing the weighting parameter α appropriately. A *timid* agent, who is reluctant to pass judgment, is expected to have a high α whereas a *self-assertive* or opinionated one is captured by a small α parameter value. The **second term** in the cost function quantifies the influence of *peer pressure* on the agent. Here, peer pressure is modeled using a quadratic cost on any opinion deviating from the mean value of others. An individualistic or independent agent is represented with a small β value. On the other hand, an agent who follows the crowd is expected to have a high-valued β parameter. The **third term**, $\gamma_i(x_i - e_i)^2$, captures the effect of the initial image e_i of an agent a_i on the final opinion x_i. A *steadfast* agent who does not change own opinion as a result of community interactions or sharing is represented by a high γ value. On the other hand, an agent who updates its opinion easily has a small γ parameter in the respective cost function. Notice that the weighting parameters α, β, γ are normalized in such a way that the factors discussed above are balanced with each other. Hence, the inherent trade-offs between the factors are captured by the cost function and the game.

The set of players or agents $\mathcal{A}$, the decision space $\mathcal{X}$, and the cost functions $J_i \ \forall i$ define together the digital trust game, $\mathcal{G}_1(\mathcal{A}, \mathcal{X}, J)$. In this noncooperative game each individual agent a_i minimizes own cost J_i by choosing own opinion (trust decision), $x_i \in \mathbb{R}$, given the opinions (trust decisions) of others, $\mathbf{x}_{-i}$, i.e.

$$x_i = \arg\min_{x_i} J_i(x_i, \mathbf{x}_{-i}). \tag{3}$$

2.1 Equilibrium Analysis

The well-known concept of **Nash equilibrium** [8] provides an appropriate solution for the digital trust game. In this context, Nash equilibrium is defined as a set of agent opinions $\mathbf{x}^*$ of a given seller (and the corresponding costs J^*), with the property that no agent has any incentive for modifying own opinion while the other agents keep theirs fixed.

The opinion of an agent given the opinions of others is uniquely determined by the best response function defined in (3). Since J_i is a polynomial strictly convex in x_i, the minimization in (3) admits a unique globally optimum solution. Consequently, the decision, x_i, of agent a_i is a unique response to any given $\mathbf{x}_{-i}$.

If the agents (players) are **symmetric** in their properties, i.e. $\alpha_i = \alpha$, $\beta_i = \beta$, and $\gamma_i = \gamma \quad \forall i$, then the Nash equilibrium solution of the digital trust game can be explicitly characterized with an analytical expression. Let $\bar{x} = \sum_i x_i$ and $\bar{e} = \sum_i e_i$. Due to strict convexity of J, it is sufficient to check the first order necessary condition for optimality

$$\frac{\partial J_i}{\partial x_i} = 0 \Rightarrow x_i^* = \left(\frac{\beta}{N-1}\sum_{j \neq i} x_j^* + \gamma e_i\right) \quad \forall i.$$

After simple algebraic manipulations, the unique Nash equilibrium of the game $\mathcal{G}_1$ is computed as

$$x_i^* = \frac{\gamma}{N-1+\beta}\left(\frac{\beta}{1-\beta}\bar{e} + (N-1)\, e_i\right) \quad \forall i.$$

Even when the agents are **not symmetric**, the uniqueness of Nash equilibrium is preserved. The best response functions of the agents can be written at the Nash equilibrium, $\mathbf{x}^*$, in matrix form $\mathbf{x}^* = \mathbf{A}\mathbf{x}^* + \mathbf{c}$, where $c_i = \gamma_i e_i \quad \forall i$ and the matrix $\mathbf{A}$ is defined accordingly. Hence, the Nash equilibrium is

$$\mathbf{x}^* = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{c},$$

where $\mathbf{I}$ is the identity matrix and $(\cdot)^{-1}$ denotes matrix inversion operation. Notice that the matrix $\mathbf{I}-\mathbf{A}$ is diagonally dominant as $A_{ii} = 1 > \sum_j |A_{ij}| = \beta_i \quad \forall i$. Therefore, it is of full rank and invertible. Consequently, the digital trust game $\mathcal{G}_1$ always has a unique Nash equilibrium solution.

2.2 Dynamics and Convergence

The agents participating in the digital trust game usually cannot reach a stable opinion in a single round. They may also change their decisions dynamically while interacting with each other, unless the system is at the Nash equilibrium. These agent dynamics can be modeled using iterative update algorithms. Parallel and Random Update Algorithms and their convergence analysis are of practical importance and provide valuable insights into the dynamical aspects of digital reputation systems.

In **Parallel Update Algorithm (PUA)**, each agent a_i updates own opinion $x_i(t)$ together (in parallel) with all other agents at the same discrete time instances $t = 1, 2, \ldots$ according to its own best response function:

$$x_i(t+1) = \frac{\beta_i}{N-1} \sum_{j \neq i} x_j(t) + \gamma_i e_i, \quad \forall i. \tag{4}$$

Therefore, PUA is also known as *synchronous update algorithm*. Algorithm 1 summarizes the steps of the PUA.

From the Perron-Frobenius theorem [9], the eigenvalues, λ of the matrix $\mathbf{A}$ satisfy

$$\min_i \beta_i \leq |\lambda| \leq \max_i \beta_i, \quad i = 1, 2, \ldots, N.$$

Hence, all of the eigenvalues of the linear system in (4) are inside the unit circle, and the PUA globally geometrically converges to the unique Nash equilibrium of the game, $\mathbf{x}^*$.

Algorithm 1 Parallel Update Algorithm (PUA)

```
Input: Individual trust values $e$, convergence threshold $\varepsilon$.
Initialize trust values $x_i(0) = e_i \ \forall i$ and time step $t = 0$.
while $\|x(t+1) - x(t)\| > \varepsilon$ do
    $t = t + 1$
    Compute $s(t) := \sum_i x_i(t)$
    for $i = 1$ to $N$ do
        Compute $x_i(t+1) = \frac{\beta_i}{N-1}(s(t) - x_i(t)) + \gamma_i e_i$.
    end for
end while
```

In many practical cases, such as in peer-to-peer (P2P) networks or e-commerce, it is not always possible to ensure that all agents update their trust decisions sequentially or synchronously in parallel. For example, some of the agents may be offline or their decision update messages may be received with delay. **Asynchronous Update Algorithm (ASU)**, where only a random subset of agents update their opinions at a given time instance, provides a realistic alternative schemes for such settings.

The ASU can be seen as a natural generalization of the PUA due to its parallel and asynchronous nature. ASU is a more suitable scheme for practical scenarios when it is difficult for the agents to synchronize their exact update instances. The ASU is defined as

$$x_i(t+1) = \begin{cases} \frac{\beta_i}{N-1} \sum_{j \neq i} x_j(t) + \gamma_i e_i & \text{,if } a_i \in \mathcal{U}(t) \\ x_i(t) & \text{,if } a_i \in \bar{\mathcal{U}}(t) \end{cases}, \tag{5}$$

where the random set $U(t)$ represents the updating agents at time t and $\bar{\mathcal{U}}(t)$ the non-updating agents. Naturally, $U(t) \cup \bar{\mathcal{U}}(t) = \mathcal{A}$. Algorithm 2 summarizes the steps of the ASU.

Algorithm 2 Asynchronous Update Algorithm (ASU)

```
Input: Individual trust values $e$, convergence threshold $\varepsilon$.
Initialize trust values $x_i(0) = e_i \ \forall i$ and time step $t = 0$.
while $\|x(t+1) - x(t)\| > \varepsilon$ do
    $t = t + 1$
    Compute $s(t) := \sum_i x_i(t)$
    for $i = 1$ to $N$ do
        if agent $i$ updates then
            Compute $x_i(t+1) = \frac{\beta_i}{N-1}(s(t) - x_i(t)) + \gamma_i e_i$.
        else
            No change in decision, $x_i(t+1) = x_i(t)$.
        end if
    end for
end while
```

The ASU converges to the unique Nash equilibrium of the trust game as it satisfies the synchronous convergence condition, which follows from the spectral radius of the matrix $|\mathbf{A}|$ being less than one, $\rho(|\mathbf{A}|) < 1$, and the box condition. Hence, global geometric convergence of ASU is established by Proposition 3.1 [10, p. 435].

For a scenario with 20 symmetric agents and parameters, $[\alpha, \beta, \gamma] = [0.2, 0.3, 0.5]$, the iterative evolution of trust under PUA is shown in Figure 1. The speed of convergence to Nash equilibrium values, which are shown with dashed lines in the figure, is geometric.

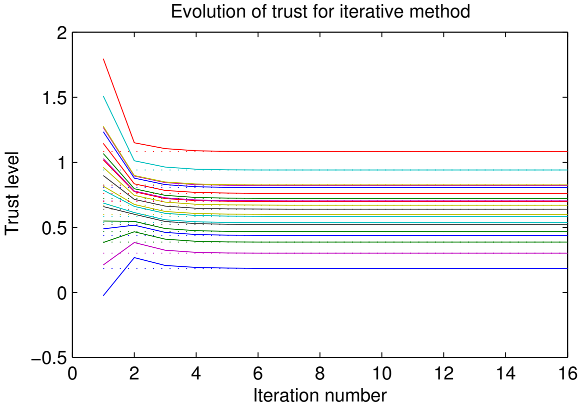

Figure 1: Evolution of trust under parallel update algorithm.

3. NUMERICAL ANALYSIS

This section presents a numerical analysis of the digital trust game using based on example scenarios, which illustrate the underlying concepts discussed such as community effects and agent properties. In each of the following scenarios, the digital trust game is played among 20 agents, who have a random initial trust level (image) of the seller, $e_i, \ i = 1, \ldots, 20$. The same initial values are used for all tests. Since the convergence properties of various update schemes are already established, the focus here is on the initial and final (Nash equilibrium) trust values of the agents, which are depicted with dark and light bars, respectively.

The first scenario studies the effects of peer pressure on agents, for example, in an online community. If the term β, which quantifies the influence of peer pressure on the agent is dominant in the cost function (2), then the agents have a strong incentive for not to deviate from the mean value of others. The parameters are $[\alpha, \beta, \gamma] = [0.2, 0.6, 0.2]$. The results show that the trust levels of all agents converge close to a common value, which can be interpreted as community opinion, as illustrated in Figure 2.

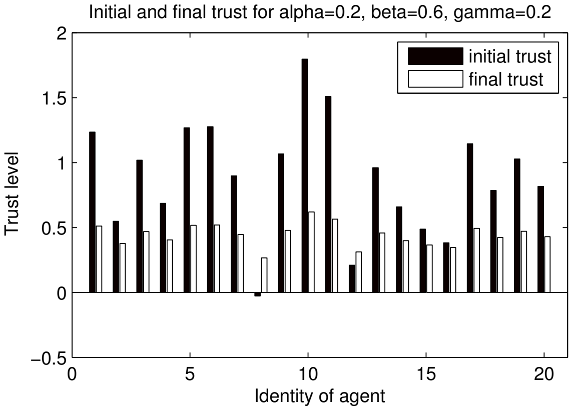

Figure 2: Initial and Nash equilibrium trust values for agents under strong peer pressure.

The second scenario investigates the case when the agents are timid, i.e. undecided or reluctant to trust or mistrust, is captured by dominant α value in the cost function. Such agents are hesitant to trust or mistrust a digital identity which causes the trust decisions converge to values close to zero (neutral opinion). The initial and final Nash equilibrium values for timid agents with the parameter set $[\alpha, \beta, \gamma] = [0.6, 0.3, 0.1]$ are depicted in Figure 3. On the other hand, if the agents are self assertive (opinionated) which is captured by having a dominant γ value, they will stick to their initial opinion on the reputation of seller. The results of a numerical analysis with self assertive agents and the parameter set $[\alpha, \beta, \gamma] = [0.1, 0.2, 0.7]$ are illustrated in the Figure 4. It is observed that there are only slight deviations in agents opinions from their initial values.

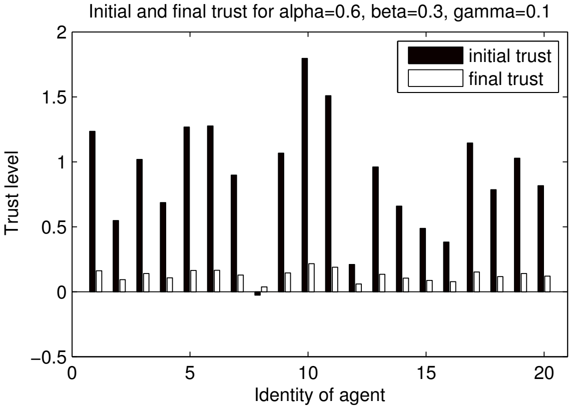

Figure 3: Initial and Nash equilibrium trust values for timid agents.

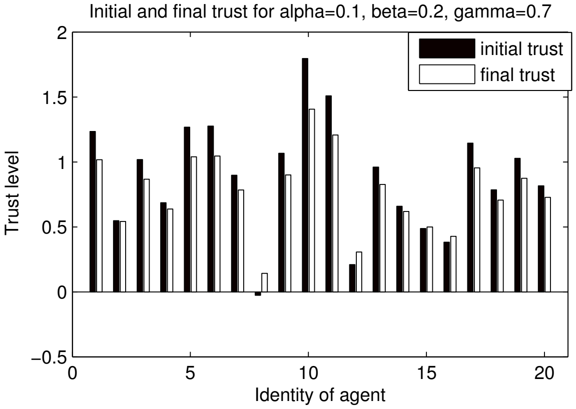

Figure 4: Initial and Nash equilibrium trust values for self assertive agents.

4. CONCLUSION

This paper presents a game theoretic model for studying evolution of trust in online communities. The quantitative model takes into account community influences and interactions between individual agents explicitly. Factors such as peer pressure and personality traits such as timidness or reluctance to pass judgment are investigated in a noncooperative game setting. The effect of various parameters on the final outcome as well as equilibrium and convergence properties of the iterative process are studied. Subsequently, the trust game and its parameters are numerically analyzed in various example scenarios.

The game theoretic framework in this paper can be seen as an initial step towards more complete and realistic models. Future research directions include an experimental study and analysis of the framework as well as further development of the game theoretic model to capture additional factors such as agent inertia.

5. REFERENCES

[1] Audun Josang, Roslan Ismail, and Colin Boyd, “A survey of trust and reputation systems for online service provision,” *Decis. Support Syst.*, vol. 43, no. 2, pp. 618–644, 2007.

[2] “eRep: Social Knowledge for e-Governance,” March 2009, FP6 European project aimed at providing theory-driven and empirically backed-up guidelines for designing reputation technologies.

[3] J. C. Ely, D. Fudenberg, and D. K. Levine, “When is Reputation Bad?,” *SSRN eLibrary*, 2004.

[4] K. Aberer and Z. Despotovic, “On reputation in game theory - application to online settings,” 2004.

[5] P. Resnick, R. J. Zeckhauser, J. Swanson, and K. Lockwood, “The Value of Reputation on eBay: A Controlled Experiment,” *SSRN eLibrary*, 2002.

[6] P. Resnick and R. Zeckhauser, “Trust among strangers in Internet transactions: Empirical analysis of eBay’s reputation system,” in *The Economics of the Internet and E-Commerce*, Michael R. Baye, Ed., vol. 11 of *Advances in Applied Microeconomics*, pp. 127–157. Elsevier Science, 2002.

[7] P. Nurmi, “A bayesian framework for online reputation systems,” in *Proc. of AICT-ICIW 2006*, Washington, DC, USA, 2006, p. 121, IEEE Computer Society.

[8] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*, SIAM, Philadelphia, PA, 2nd edition, 1999.

[9] R. Horn and C.R. Johnson, *Matrix Analysis*, New York, NY: Cambridge University Press, 1985.

[10] D. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Compuation: Numerical Methods*, Prentice Hall, Upper Saddle River, NJ, 1989.