Catch me, if you can: 
The long path from reputation to trust

Karl Aberer
EPFL

joint work with
Zoran Despotovic, Le Hung Vu,
Thanasis Papaiaonnou
MOTIVATION
Trust in Online Systems

- No trusted authority
- Selfish behavior
- Examples
  - Peer-to-peer system
  - Social networks
  - Ecommerce sites
  - Internet gaming
Selfish Behavior

• Example P2P: to conserve
  – Own bandwidth
  – Own CPU cycles
  – Freeriding

• To exploit
  – Greedily consume other’s resources
  – ... before others can use that opportunity
Example of Selfishness

Freeriding in Gnutella [Adar, Hubermann 2000]

Cooperation needs to be built: *but how?*
Cooperation Based on Trust

• **Trust:** the extent to which a peer believes the other peer will cooperate
  – Gain is to be shared with the other peer
  – But the peer is exposed to a risk of loss
Cooperation Based on Trust

• **Trust**: the extent to which a peer believes the other peer will cooperate
  – Gain is to be shared with the other peer
  – But the peer is exposed to a risk of loss
Cooperation Based on Trust

• **Trust**: the extent to which a peer believes the other peer will cooperate
  – Gain is to be shared with the other peer
  – But the peer is exposed to a risk of loss
Cooperation Based on Trust

- **Trust model**: the way in which the above belief is inferred
  - Inference could be: if Bob cooperates with Alice then Bob will also cooperate with me with high probability
Cooperation Based on Trust

• **Trust model**: the way in which the above belief is inferred
  
  – Inference could be: if B cooperates with C then B will also cooperate with A with high probability
Cooperation Based on Trust

• **Trust model**: the way in which the above belief is inferred
  – Inference could be: if B cooperates with C then B will also cooperate with A with high probability
Cooperation Based on Trust

• **Trust model**: the way in which the above belief is inferred
  – Inference could be: if B cooperates with A then B will also cooperate with C with high probability

• **Reputation-based trust**: Inference based on the collective earlier actions of a peer:

  it’s reputation
Does it really matter?

• Example: eBay
  – eBay reputation profiles predictive of future performance [Resnick et al., 2002]
  – Prices positively correlated with the feedback [Melnik and Alm, 2002]
Overview

MOTIVATION

REPUTATION-BASED TRUST IN ONLINE COMMUNITIES

TRUST WITH RATIONAL PEERS

TRUST AND IDENTITY

CONCLUSIONS
REPUTATION-BASED TRUST IN ONLINE COMMUNITIES
The Path Towards Trust

Problem

Selfish Behavior

Countermeasure

Reputation
On-line Reputation

• Common characteristics
  – open community: interactions with different parties
  – rarely repeated interactions with the same party
    • Therefore need for recommendations of others
  – possibility to misbehave and misreport
The Path Towards Trust

Problem

- Selfish Behavior
- Manipulation of Reports

Countermeasure

Reputation
Model

- Peers provide services to other peers
  - Know other peers they interacted with
  - Store feedback $w_i$ on interactions with peer $i$
  - Feedback on service and reporting
  - A social network (trust graph)
Social Networking Approach

• Compute global trust values $t_j$ for peer $j$
  – By aggregating feedbacks
  – Weighting by the trust in the recommender
  – Similar to PageRank computation

$$t_j = \sum_{i \in \text{in}(j)} w_i \frac{t_i}{\sum_{k \in \text{in}(i)} t_k}$$

[Xiong, Liu TKDE 2004]
Social Networking Assessment

• Problems solved
  1. Automatic aggregation of reputation data
  2. Does not require central authority
  3. Possibility of misreporting considered

• Shortcomings
  – Costly (global) computation
  – Trust values have no further meaning but ranking
  – No distinction between propensity to provide poor service and to misreport
Probabilistic Estimation Approach

• Assume probabilistic peer behavior, eg.
  – $P[\text{peer } j \text{ performs service honestly}] = \theta_j$
  – $P[\text{peer } k \text{ lies when reporting}] = l_k$

• Probability of report $y_k$ on peer $j$ from peer $k$

$$P[Y_k = y_k] = \begin{cases} 
  l_k (1-\theta_j) + (1-l_k)\theta_j & \text{if } y_k = 1 \\
  l_k \theta_j + (1-l_k)(1-\theta_j) & \text{if } y_k = 0
\end{cases}$$

[Despotovic, Aberer Journal of Computer Networks 2004]
Probabilistic Estimation Approach

• Maximum Likelihood Estimation
  – Determine \( l_k \) by checking reports on own performance
  – Collect all reports \( y_1, y_2, ..., y_n \) on peer \( j \)
  – Select \( \theta_j \) that maximizes the probability to obtain the observed reports

\[
L(\theta_j) = P [Y_1 = y_1] P [Y_2 = y_2] \cdots P [Y_n = y_n]
\]
Comparative Evaluation

• Precision in correctly assessing misbehavior

Maximum Likelihood Estimation

Social Networking
Comparing eBay and P2P

• Differences
  1. no centralized authority to manage and verify reputation information
  2. possibility to manipulate reports of others

• Question
  – Which methods do exist for efficiently and automatically managing reputation data in the absence of any centralized infrastructure?
Implementation Issues

• Variant 1: use trust graph as *unstructured overlay* network
  – Simple maintenance, high cost of retrieval of reports

• Variant 2: use a separate *structured overlay* network
  – Complex maintenance, efficient retrieval

• Additional problem: manipulation of stored reports and messages
  – Use replication of reputation data
Probabilistic Estimation Assessment

• Problems solved
  – Efficient trust computation
  – Trust values predict probability of cooperative behavior
  – Distinction between propensity to provide poor service and to misreport

• Shortcomings
  – Reputation-based trust detects malicious and selfish behaviors (signaling)
  – ... but does not consider rationality of agents (sanctioning – fostering honest behavior)
The Path Towards Trust

Problem

- Selfish Behavior
- Manipulation of Reports
- Strategic Behavior

Countermeasure

- Reputation
- Reputation of recommenders

Computational trust models for signalling
REPUTATION-BASED TRUST WITH RATIONAL PEERS
A Rational Agent

Reputation

Detection threshold
Rationality

• Actions of peers have associated utility
• Example: prisoners dilemma

• Main insights
  – One shot game: no cooperation (Nash)
  – Repeated game: cooperation – tit-for-tat (Axelrod)
• Common explanation of the concept of trust
Extended Model

• Model
  – Consumers provide feedback $w_i$ on peer $i$’s service
  – Peer $i$ has incentive to cheat:
    legal gain $u_i <$ illegitimate gain $u_i + v_i$

• Peer behavior
  – Honest peers never cheat
  – Malicious peers cheat probabilistically
  – Rational peers optimize their utility

[Hung, Aberer WI 2008, TAAS 2010]
Sanctioning

If \( t = 1 \) and \( w = 0 \) or \( t = 0 \) and \( w = 1 \):

3 Peer \( i \) is cheating!!!

Most recent feedback

Evaluate reliability \( t \)

of feedback \( w \)

using

computational trust model

If peer \( i \) has less than \( k \) cheating detections: ok!

Severe sanctioning mechanism!
Computational Trust Model

• may use several information sources
  – past rating behavior of the rater, past performance of sellers, trusted sources, own belief on environment’s vulnerability, relations between peers

• may use variety of statistical models/heuristics
  – probabilistic approaches, collaborative filtering, similarity of rater and own’ experience, clustering of ratings to isolate dishonest rater

• Accuracy measure (known to peers)

\[ P[est+ | real-] < \varepsilon, \ P[est - | real+] < \varepsilon \]
Trust Accuracy vs. Cooperation

**Theorem:** if computational trust model sufficiently accurate and gains are bounded rational peers cooperate in all but the last $\Delta$ transactions.

**Bounded gains** ($u_* < u < u^*$, $v_* < v < v^*$)

$$\Delta = \max\{1, \left\lceil \frac{\ln[1 - \frac{v^* \varepsilon^k}{u^*(1 - \varepsilon + k - \varepsilon^k)}]}{\ln(1 - \varepsilon^k)} \right\rceil \}$$

$$\varepsilon < \varepsilon_{\text{max}}(k) = \frac{1}{1 + \sqrt[k]{1 + \frac{v^*}{u_*}}}$$
Incentive to Leave ($\Delta$)

Example:
For $\varepsilon < 0.2$ and $k=1, 2$ peer will leave for the last 2 transactions
Emergence of Cooperation

- Cooperation is enforced if peers stay infinitely or long enough given $\varepsilon$ sufficiently small
  - resilient against rating manipulation
  - malicious peers are eliminated

Example: For $\varepsilon = 0.2$ and $k=10$ peer will be accidentally blacklisted after $2^{22}$ transactions ...

... whereas a cheater will be eliminated after $2^3$ cheats
Accuracy-cost Trade-off

- Presence of high quality dishonesty detector may prevent rating manipulation by rational sellers
- High accuracy implies higher implementation cost
- Combine two computational trust models
  - expensive/accurate dishonesty detector with probability $c$
  - trust the rating with probability $1-c$
- Result
  - With very low $c$ rational peers still cooperate
Sanctioning: Assessment

• Considering rational behavior
  – Enforces cooperation of rational peers
  – Eliminates malicious peers
  – Permits to optimize computational cost for dishonesty detection

• Shortcomings
  – Requires shared, secure storage for reputation data
  – Requires stable and global identities (like all other reputation-based trust mechanisms)
The Path Towards Trust

**Problem**
- Selfish Behavior
- Manipulation of Reports
- Strategic Behavior
- Whitewashing

**Countermeasure**
- Reputation
- Reputation of recommenders
- Sanctioning
TRUST AND IDENTITY
Key Problem

The problem of trust is inherently linked to stable identification
How to prevent whitewashing?

• Remember: clients on eBay are willing to pay higher prices for reliable seller
  – 8.1% according to a study of [Resnick 2006]

• *Can this help?*

• Idea: sellers that stay longer in the system ask for higher prices
  – Makes it unattractive to leave the system for whitewashing

• *Does it work?*
Identity Premium

\[ P(L) = u (1 - \Phi) + f(L) \]

- \( L \) lifetime of seller
- \( f(0) = 0, f \) monotonically increasing
- \( 0 < \Phi < 1, \) initial price below original price \( u \)
Cooperation Enforcement

• **Theorem:** If the identity cost is sufficiently small there exists an identity premium function such that a rational provider will cooperate in every but the last interaction.

• Bound on identity cost and premium depends on
  – Accuracy of dishonesty detector
  – Potential cheating gain
  – Initial price
THEOREM 1. Given the provider selection protocol $S_k = \langle R, k \rangle$ where the dishonesty detector $R$ has the misclassification errors $\alpha, \beta$ upper-bounded by $\varepsilon < 0.5$.

Consider any rational provider with $N$ services to sell. Let $u_* \leq u \leq u^*, i = 1, ..., N$ be the original prices of the services sold by the provider in the $i$-th transaction. Suppose that the pricing scheme $P(\phi, f)$ is used, it follows that:

(i) If the identity cost $\xi$ is small, the following identity premium ensures that cooperation is always the best response strategy of the provider in any transaction $i = 1, ..., N - 1$, for any $0 < \phi_i < 1$:

$$f(L) = \sum_{i=1}^{L} \lambda^{L-i} (\lambda u(1 - \phi_i) - \xi/\gamma) \text{ for } L > 0$$

(ii) For $\lambda \neq 1$ and providers sell services of the same standard price $u = 1, \phi_i = \phi, i = 1, ..., N$, if the identity cost $\xi < \xi_0 = \gamma \lambda (1 - \phi)$, the following identity premium function is sufficient to enforce cooperation for a provider in selling every but the last service:

$$f(L) = ((1 - \phi)\lambda - \xi/\gamma) \frac{1 - \lambda^L}{1 - \lambda} \text{ for } L > 0$$

For $\lambda = 1$, the identity premium function becomes:

$$f(L) = L (1 - \phi - \xi/\gamma) \text{ for } L > 0$$

(iii) Let $N_h$ be the number of transactions a fully cooperative (honest) provider can participate till it is mistakenly blacklisted, and let $N_c$ be the number of bad transactions an intentionally malicious provider can benefit from defecting until eliminated from the system, respectively. We have $E[N_h] > 1/\varepsilon^k$ and $E[N_c] < 1/(1 - \varepsilon)^k$.

The results (i,ii,iii) hold even in presence of strategic manipulation of ratings by agents.
Example

- original price $u=1$
- cheating gain 50% of original price
- Initial price $0.5$ ($\Phi = 0.5$)

Limit price above original price: Acceptable to buyers?

Inefficiency?

Limit price below original price: Acceptable to sellers?
Eliminating Inefficiency

• If the cheating gain is sufficiently small and the dishonesty detector is sufficiently accurate the price remains bounded and can approach any limit price by properly choosing $\Phi$
  – Thus for a finite number of services this inefficiency can be eliminated
  – For infinite number it can be kept extremely small (order of provisioning a few services)

• Rationale to accept the scheme
  – Providers: without premium no trade at all
  – Consumers: no risk if providers are rational (provably)
CONCLUSIONS
The Path Towards Trust

Problem
- Selfish Behavior
- Manipulation of Reports
- Strategic Behavior
- Whitewashing

Countermeasure
- Reputation
- Reputation of recommenders
- Sanctioning
- Identity Premium
From Closed to Open Trust

• So far trust managed in a closed system, but
• Multiple trust systems
  – Transferrable identity
  – Aggregate reputation from multiple sources
• Open Web
  – Semantic Web: content-based → entities
    • All Web content on Tim-Berners Lee
  – Social Web: friend-to-friend authentication
    • Social networks as computing substrate
  – Sensor Web: link the web to physical reality
    • Soft biometrics, GPS, ...
New Opportunities for Trust Systems

• Semantic, Social and Sensor Web not only help to build trust but are also in need of trust
  – Semantic Web: is the content trustworthy?
  – Social Web: are my social connections trustworthy?
  – SensorWeb: is the measured data reliable?
Thank you for your attention!

... and we will continue to keep you busy for a while
Acknowledgements

• Work sponsored by
  – Swiss National Science Foundation
  – EU FP 6, projects DIP, Nepomuk
  – NTT Docomo EuroLabs, Munich
Main Publications


