Trust Management
from Security to Reputation Mechanisms

Bin Yu
Department of Computer Science
North Carolina State University

Trust Management

Trust management

- any two parties can authenticate each other to the extent that they are willing to undertake the transaction.
  - Traditional face-to-face transactions
  - Electronic commerce.

Trust management in security

- Most are based on public key “certificate” in which a trusted third party or any party signs a special message certifying the identity associated with a public key.
  - PGP-style web of trust
  - X.509-style certifying authority trees

PGP system

PGP – Pretty Good Privacy

- Primarily for encrypting email messages using public key cryptography.

A user generates a (*PublicKey, SecretKey*) pair that is associated with his unique ID

- Usually an ID is of the form (*Name, EmailAddress*).
- A public key contains an ID, a public key, and a timestamp of when the key pair was created.

Example

- If user A has a good copy of user B's public-key, e.g., a copy has not been tampered with since B generated,
- Then A can sign this copy and pass it on to user C.
- A thus acts as an *introducer* of B to C.

PGP Web of Trust

Each user must tell the PGP system who he trusts as introducers and must certify the introducers' public-key with his own secret key.

A user may specify the *degree of trust* that he has in each introducer.

- Unknow, untrusted, marginally trusted, or completely trusted.

Trust is not transitive in PGP

- A fully trusts B as an introducer and B fully trusts C do not automatically imply anything about A's degree of trust in C.

X.509

X.509 certificates contain more information than PGP certificates

- The names of the signature schemes
- The time interval in which they are valid.

X.509 differs from PGP in its level of centralization of information

- Anyone may sign public-key and act as an introducer in PGP.
- The X.509 postulates that everyone will obtain certificates from an official *certifying authority* (CA).
  - An authority trusted by one or more users to create and assign public key certificates.

X.509 Certifying Authority Tree

If A and B have both been certified by the same CA

- The directory server can just send B's certificate to A, who can verify its validity using the public key of this common CA.

If A and B have not been directly certified by a common CA

- The directory server must create a certification path from A to B.
- A list of the form CA_1, $cert_1$, CA_2, $cert_2$, …, CA_n, $cert_n$, where $cert_i$, $1=<i<n$, is a certificate of CA_{i+1}, that has been signed by CA_i and $cert_n$ is a certificate of B.
- In order to use this path, A must know the public key of CA_1, the first authority in the path.

Trust and Security

Trust management in security

- PGP-style web of trust
- X.509-style certifying authority trees

Problems of trust management in security

- Trust is more than creating, acquiring and distributing certificates.
  - A party is authenticated and authorized, but this does not ensure that it exercises in a way that is expected.
- Trusted Third Parties are not always available in an open and dynamic environment.
  - X.509, especially a multiple connected one, is much more expensive to build up.

Reputation Mechanisms

Online reputation system

- Collects, distributes, and aggregate feedback about participants' past behavior.
- Help people decide whom to trust, and deter participation by those who are unskilled or dishonest.

Examples

- OnSale
  - Allows users to rate sellers.
  - The overall reputation of a seller is the average of the ratings obtained from users.
- Ebay
  - Sellers receive feedback (-1, 0, 1) in each auction.
  - Reputation of a seller is calculated as the sum of its ratings over the last six months.

Challenges

Significant challenges for reputation systems

Eliciting

- People may not bother to provide feedback at all.
  - When a trade is completed, there is little incentive to spend another few minutes filling out a form.
- It is hard to assure honest reports.
  - A group of people might collaborate and rate each other, artificially inflating their reputations.

Distributing

- Name changes
  - They can choose another pseudonym, effectively erasing prior feedback.

Aggregating

- The simple numerical ratings fail to convey important information of online transactions.

Distributed Reputation Management

Distributed reputation management

- Decentralizes the sources of reputation.
- A promising approach for achieving robustness in the presence of potential dishonest participants and privacy concerns.

Our approach

- adjusts the ratings of agents based on their observations as well as the testimony from others
- helps agents (users) avoid interaction with non-cooperative (malicious) participants.

Prisoner's Dilemma

Security itself does not promote cooperation

- Cooperation is rational only if a large number of interactions are expected in the future.
- E.g., Axelrod on the Prisoners' Dilemma.

Reputation mechanisms promote rational cooperation in large and dynamic distributed systems

- More than iterated prisoners' dilemma (not repeatedly interact, various initial decisions).
- The aggregate rewards and penalties from a society are greater than from an individual.

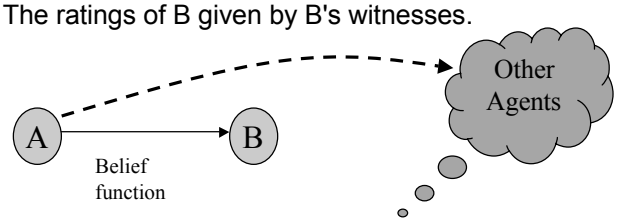
Framework

Each participant (user) has a personal agent and the agents assist their users in

- Evaluating the services and referrals provided by others.
- Deciding whom to contact for a service.

Agent A rates agent B based on

- Its direct observations of B.
- The ratings of B given by B's witnesses.



12

Research Challenges

Local ratings

- Captures the ratings over the last several interactions,
- Then converts the ratings into belief functions.

Witnesses

- A process of referrals, in which each agent being queried offering referrals to other agents.
- This leads to a focused search that does not send irrelevant messages to other agents.

Testimonies

- Our approach includes the necessary representation and reasoning through which testimonies can be combined in a principled manner.

Dempster-Shafer Theory

Frame of discernment

- The given set of propositions
- Here, T and ¬T: whether to trust the other agent.

Basic probability assignment

- If Θ is a frame of discernment, then *m*: 2^{Θ} [0,1] is a *basic probability assignment* (*bpa*) whenever
  (1) m(∅) = 0, where ∅ is an empty set, and
  (2) $\sum_{A \subseteq \Theta} m(A) = 1$, where A is a subset of Θ.

Belief function of a set

- Sum basic probability assignments over all subsets of the set.

Belief Rating

Suppose

- Agent A_i has the latest h responses from agent A_j, S_j = $\{s_{j1}, s_{j2}, \ldots, s_{jh}\}$.
- The quality of service (QoS) $s_{jk} \in \{0.0, 0.1, \ldots, 1.0\}$.
- Two thresholds α_i, and β_i.

Then agent A_i can get the *bpa* toward agent A_j

$$m(\{T\}) = \sum_{x_k = \alpha_i}^{1} f(x_k)$$

$$m(\{\neg T\}) = \sum_{0}^{x_k = \beta_i} f(x_k)$$

$$m(\{T, \neg T\}) = 1 - m(\{T\}) - m(\{\neg T\})$$

Where $f(x_k)$ denotes the probability that a particular quality of service (QoS) x_k happens.

15

Rules of Combination

Suppose

- Bel_1 and Bel_2 are belief functions over the same frame Θ, with *bpa* m_1 and m_2, and focal elements $A_1, \ldots, A_k$, and $B_1, \ldots, B_l$, respectively.
- $$\sum_{i,j,A_i \cap B_j} m_1(A_i) m_2(B_j) < 1$$

Then the function $m : 2^{\Theta} \rightarrow [0,1]$ defined by

- $m(\emptyset) = 0$, and
- $$m(A) = \frac{\sum_{i,j,A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - \sum_{i,j,A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)}$$

For all non-empty $A \subseteq \Theta$, m is a bpa.

Examples

Given two belief functions,

$m_1(\{T\}) = 0.8$, $m_1(\{\neg T\}) = 0$, $m_1(\{T, \neg T\}) = 0.2$
$m_2(\{T\}) = 0.9$, $m_2(\{\neg T\}) = 0$, $m_2(\{T, \neg T\}) = 0.1$

Then
$m_{12}(\{T\}) = m_1(\{T\})\, m_2(\{T\}) + m_1(\{T\})\, m_2(\{T, \neg T\}) + m_2(\{T\})\, m_1(\{T, \neg T\})$
$= 0.72 + 0.18 + 0.08 = 0.98$
$m_{12}(\{\neg T\}) = 0$
$m_{12}(\{T, \neg T\}) = 0.02$

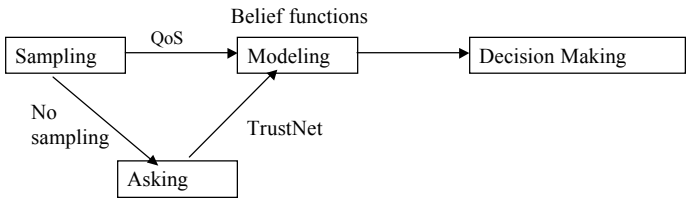
Local and Total Belief

Local belief

- $\neg$ - From direct interactions and can be propagated to others upon request.

Total Belief

- $\neg$ - Combines the local belief (if any) with testimonies received from any witnesses.



Incorporating Multiple Testimonies

A TrustNet is a directed graph TN(A_r, A_g, □, R), where □ is a finite set of agents {A_1 , …, A_N } and R is a set of referrals {r_1 , …, r_n }.

Given a set of testimonies □={w_1 , w_2 , … , w_L}, agent A_r will update its total belief rating of agent A_g as follows

$$\square_{A_r} = \square_{A_r} \oplus \square_{w_1} \oplus \ldots \oplus w_{w_L}$$

19

Experimental Setup

Each agent has

- An interest vector and an expertise vector
- A set of neighbor models

Acquaintance Models include

- other agents' expertise (ability to produce correct domain answers),
- sociability (ability to produce accurate referrals), and
- samplings of recent interactions.

Metrics

Metric 1: The average reputation of agent A_i

$$B_{A_i} = \square_{w_1} \oplus \square_{w_2} \ldots \oplus \square_{w_L}$$

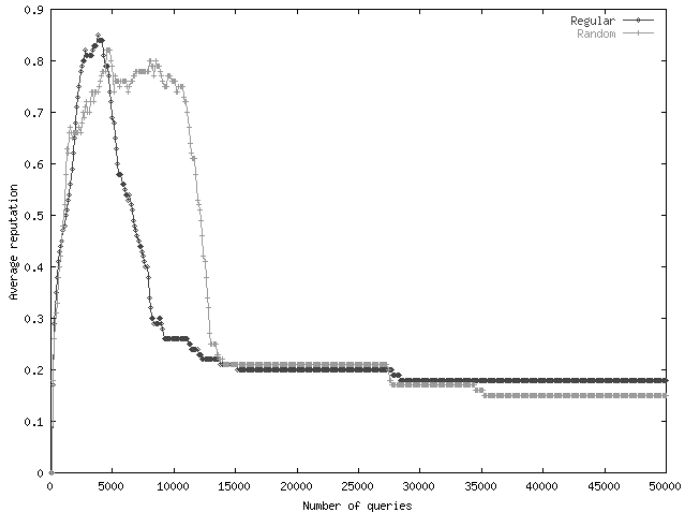
$$\square(A_i) = B_{A_i}(\{T_{A_i}\}) \square B_{A_i}(\{\square T_{A_i}\})$$

Metric 2: The average reputation of all agents

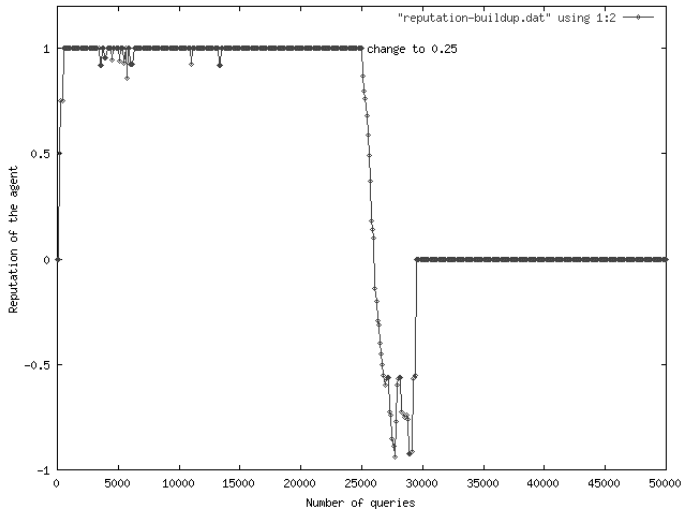
$$\square = \frac{1}{N} \square_{i=1}^{N} \square(A_i)$$

21

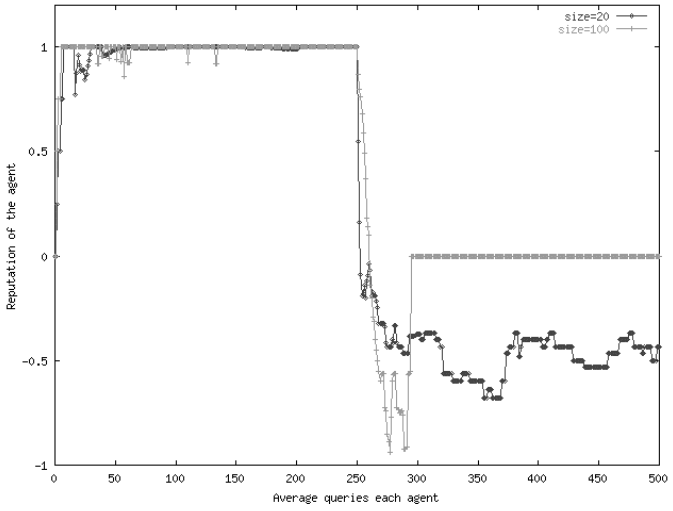
Bootstrapping



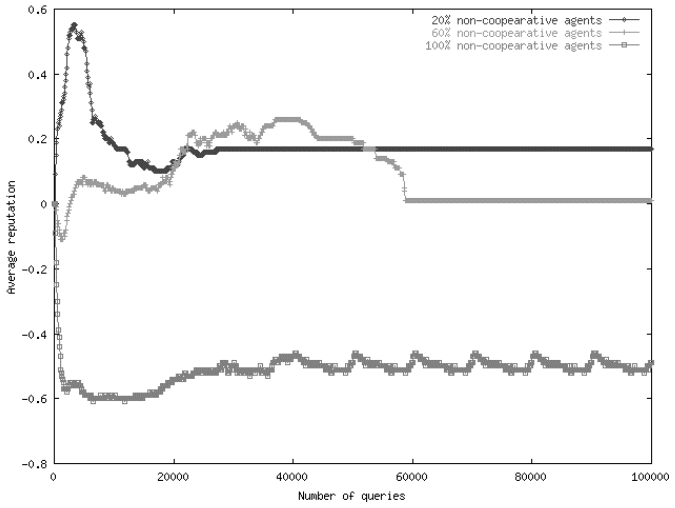
Reputation Buildup



Community Size



Ratio of Non-Cooperative Agents



Conclusion

Distributed reputation management

- _ provides an automatic, and efficient approach to detect non-cooperative (malicious) agents in an open, and dynamic environment.
- _ Leads to a decentralized society in which agents help each other weed out undesirable players.
- _ Complements cryptographic techniques in security such as passwords, public keys, and digital certificates.

Future work:

- _ Detection of deception in testimony propagation.
- _ Mechanism design, incentive of help, evolution of (in)direct reciprocity.

26

References

Matt Blaze, Joan Feigenbaum and Jack Lacy, Decentralized Trust Management, In *Proceedings of IEEE Conference on Security and Privacy*, pages 164-173, 1996

Paul Resnick, Richard Zeckhauser, Eric Friedman and Ko Kuwabara, Reputation Systems, *Communications of the ACM*, 43(12), 45-48, 2000

Chrysanthos Dellarocas, Online Reputation Mechanisms, in *Practical Manual of Internet Computing (edited by Munindar P. Singh)*, to appear

Bin Yu and Munindar P. Singh, A Social Mechanism of Reputation management in electronic communities, In *Proceedings of Fourth International Workshop on Cooperative Information Agents*, pages 154-165, 2000

Bin Yu and Munindar P. Singh, An Evidential Model of Distributed Reputation Management, *Proceedings of First International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 294-301, 2002

Bin Yu and Munindar P. Singh, Distributed Reputation Management for Electronic Commerce, *Computational Intelligence*, 18(4), 294-301, 2002