# An Integrated trust and reputation model for open multi-agent systems

## A paper by Trung Dong Huynh, Nicholas R. Jennings & Nigel R. Shadbolt (2006)

Jaspreet Singh & Daniël Stekelenburg

Universiteit Utrecht

[Faculty of Science
Information and Computing Sciences]

# Overview

1. Terminology
2. The FIRE Model
3. Results
4. Conclusions

# .. an open MAS?

"…systems in which agents can freely join and leave at any time and where the agents are owned by various stakeholders with different aims and objectives."

# .. an open MAS?

> "…systems in which agents can freely join and leave at any time and where the agents are owned by various stakeholders with different aims and objectives."

This causes some uncertainties:

1. Agents tend to be self-interested and may be unreliable
2. No agent can know everything about the environment
3. No central authority can control everything

# Sources of trust/reputation

| Source | Type |
|---|---|
| Direct experience | Interaction trust |
| Witness experience | Witness reputation |
| Role-bases rules | Role-based trust |
| Third-party references | Certified reputation |



# Fire

Uses all four sources of information

Works, based on the following assumptions:

- Agents are willing to share their experiences with others (as witnesses or as referees)
- Agents are honest in exchanging information with one another.

# Fire

Uses all four sources of information

Works, based on the following assumptions:

- Agents are willing to share their experiences with others (as witnesses or as referees)
- Agents are honest in exchanging information with one another.

So… we do not consider the problem of lying and inaccuracy.

# How to quantify trust/reputation? - The old way

Just take the average of all the ratings.

# How to quantify trust/reputation? - The old way

Just take the average of all the ratings.

However… these ratings are not equally relevant:

- Older ratings might not be as relevant as new ones
- Some ratings are more credible than other depending on the source

So in what other way can we quantify trust?



# How to quantify trust? - The FIRE way

Every rating is a tuple $r = (a, b, c, i, v)$.

Where $a$ and $b$ are the agents participating in transaction $i$. Value $v \in [-1, +1]$ is the rating given by agent $a$ to agent $b$ regarding regarding topic $c$ (e.g. quality, honesty).

These ratings are stored in the agent's local database.

# How to quantify trust? - The FIRE way

Every rating is a tuple $r = (a, b, c, i, v)$.

Where $a$ and $b$ are the agents participating in transaction $i$. Value $v \in [-1, +1]$ is the rating given by agent $a$ to agent $b$ regarding regarding topic $c$ (e.g. quality, honesty).

These ratings are stored in the agent's local database.

Since ratings become outdated over time, an agent only stores the latest $H$ transactions it gave to other agents.



# How to quantify trust? - Trust value $\mathcal{T}_K$

Use a rating weight function (reliability function) $\omega_K$ for every type of trust, where $K \in \{I, R, W, C\}$.

# How to quantify trust? - Trust value $\mathcal{T}_K$

Use a rating weight function (reliability function) $\omega_K$ for every type of trust, where $K \in \{I, R, W, C\}$.

This gives us:

$$\mathcal{T}_K(a,b,c) = \frac{\sum_{r_i \in \mathcal{R}_K(a,b,c)} \omega_K(r_i) \cdot v_i}{\sum_{r_i \in \mathcal{R}_K(a,b,c)} \omega_K(r_i)} \tag{1}$$

- $\mathcal{T}_K(a,b,c)$ is the trust value of agent $a$ towards agent $b$ on topic $c$, regarding $K$.
- $\mathcal{R}_K(a,b,c)$ are the ratings collected on $K$.
- $\mathcal{T}_K(a,b,c) \in [-1,+1]$

# How to quantify trust? - Reliability

- We now have a trust value $\mathcal{T}_K$
- How reliable is $\mathcal{T}_K$?

# How to quantify trust? - Reliability

- We now have a trust value $\mathcal{T}_K$
- How reliable is $\mathcal{T}_K$?

- We need a value to express how reliable the calculated trust value $\mathcal{T}_K$ is!

# How to express reliability?

- We know how to calculate how reliable each individual rating is: $\omega_K$
- We use this to express:
  - Rating reliability $\rho_{RK}$: The total reliability of the individual ratings.
  - Deviation reliability $\rho_{DK}$: The higher the variance in the ratings is, the more volatile the agent is likely to fulfilling its agreements.

# How to express reliability? - Rating reliability

- The total reliability of the individual ratings. → The sum of reliability of the individual ratings.

$$\rho_{RK}(a,b,c) = 1 - exp\Big( -\gamma_K \Big( \sum\nolimits_{r_i \in \mathcal{R}_K(a,b,c)} \omega_K(r_i) \Big) \Big) \quad (2)$$

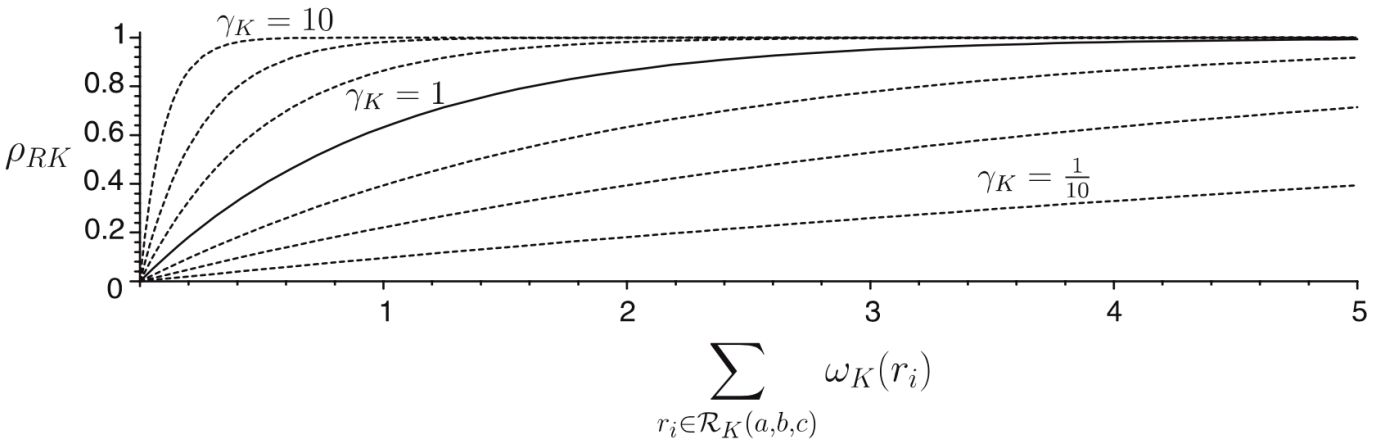


Figure 1: Rating reliability function

# How to express reliability? - Deviation reliability

- The higher the variability in the ratings is, the more volatile the agent is likely to fulfilling its agreements.
- $\rightarrow$ The higher the variability the lower the deviation reliability is.

$$\rho_{DK}(a,b,c) = 1 - \frac{1}{2} \cdot \frac{\sum_{r_i \in \mathcal{R}_K(a,b,c)} \omega_K(r_i) \cdot |v_i - \mathcal{T}_K(a,b,c)|}{\sum_{r_i \in \mathcal{R}_K(a,b,c)} \omega_K(r_i)} \quad (3)$$



# How to express reliability?

- Now we know how to calculate both the rating reliability $\rho_{RK}$ and deviation reliability $\rho_{DK}$.
- We combine both values and get a single value for the reliability of $\mathcal{T}$:

$$\rho_K(a, b, c) = \rho_{RK}(a, b, c) \cdot \rho_{DK}(a, b, c) \tag{4}$$



# Interaction trust

- Is built from the direct experiences of an agent and models the direct interactions between two agents.
- The reliability $\omega_I(r_i)$ of a single interaction is determined by its recency:

$$\omega_I(r_i) = exp\Big( - \frac{\Delta t(r_i)}{\lambda} \Big) \tag{5}$$

- $\Delta t(r_i)$ is the difference in time between now and the time when $r_i$ was recorded.
- $\lambda$ is the recency scaling factor.

# Interaction trust

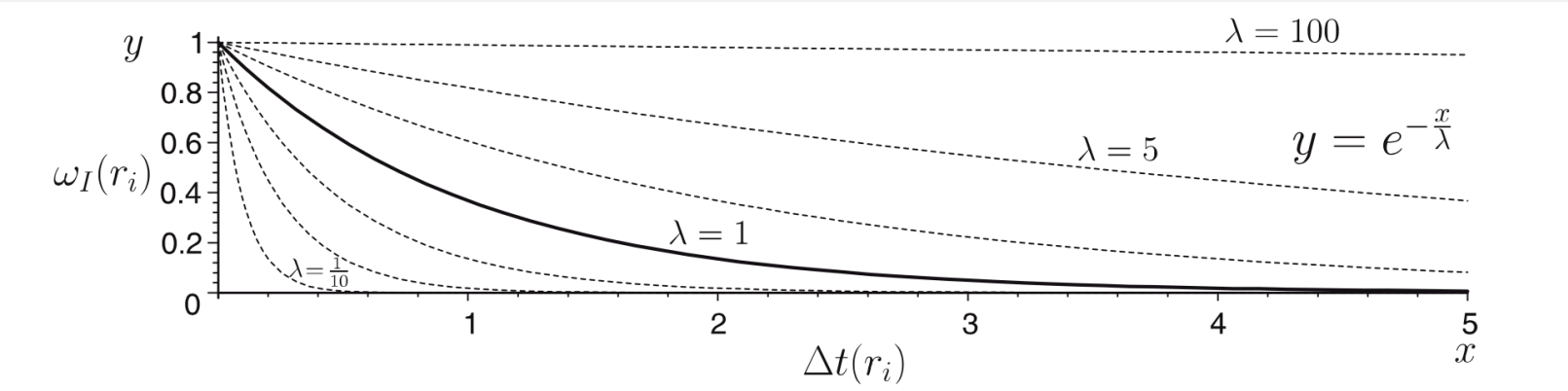


Figure 2: Behavior of the weight function $\omega_I(r_i)$.

# Role-based trust

- Models trust resulting from role-based relations.
- For example: provider-consumer relationship.
- The reliability $\omega_R(r_i)$ of a single interaction is determined by a set of rules:

$$rule = (role_a, role_b, c, e, v) \tag{6}$$

  - $v$ is the expected performance.
  - $e$ is the amount of influence this rule has on the total value.

- $\omega_R(r_i) = e_i$

# Witness reputation

- Is built on observations on the agents behavior by other agents.
- Need to find other agents that have interacted with $b$.
- This might be problematic in large environment:
  - Limited resources available;
  - Need to find these witnesses in reasonable time.
- Once all the ratings have been collected, the weight is determined by $\omega_W(ri) = omega_W(ri)$.
- Based on the idea of referrals.

# Witness reputation

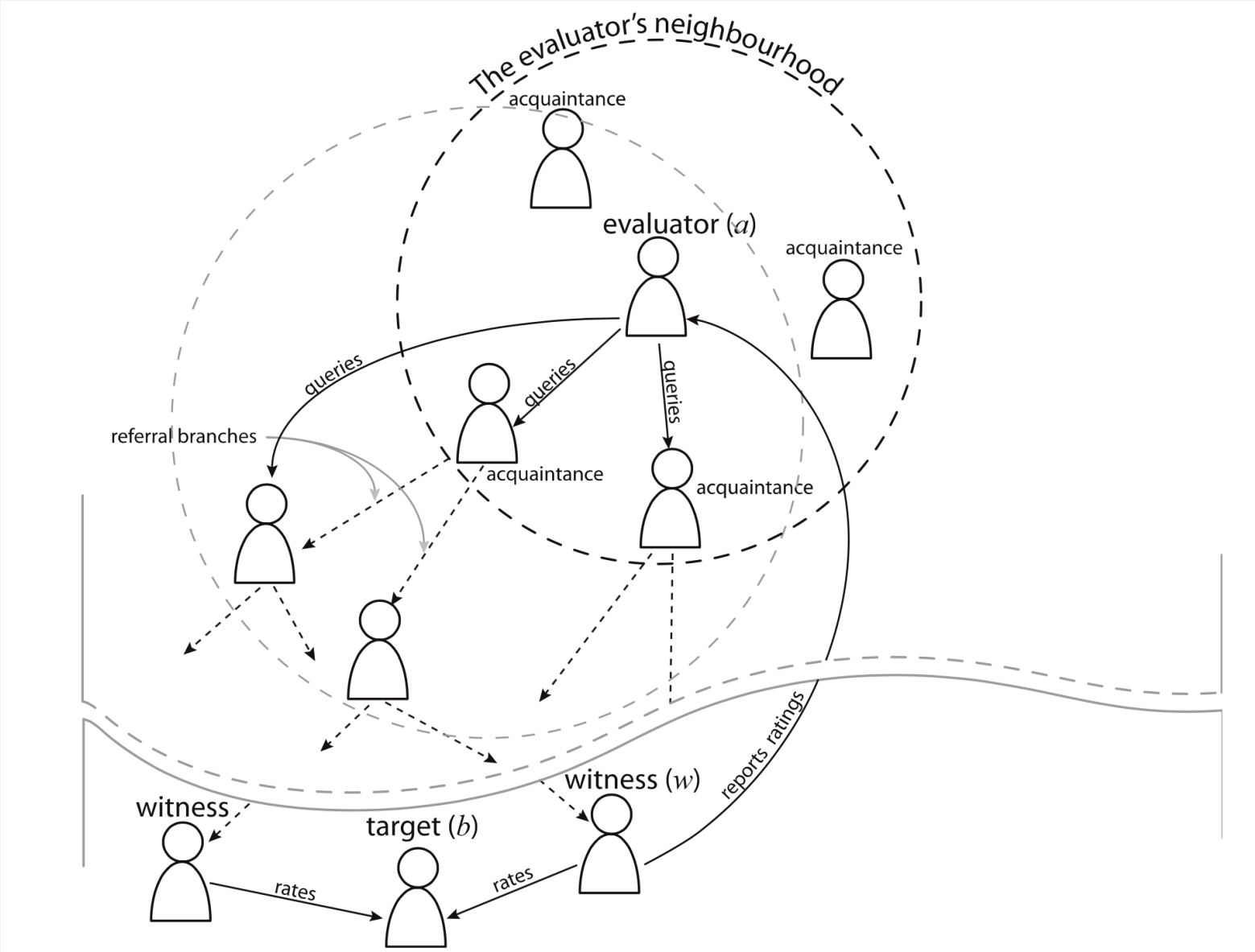


Figure 3: How to find witnesses

# Certified reputation

- Is built from ratings from certified references given by referees.
- Stored by the agent itself and chooses which ratings to present.

# Certified reputation

- Is built from ratings from certified references given by referees.
- Stored by the agent itself and chooses which ratings to present.

- After every transaction, $b$ asks $a$ to give a certified rating.
- When $a$ contacts $b$, it asks $b$ for the te certified references.
- Since the ratings are from direct interactions, $\omega_C(r_i) = \omega_I(r_i)$.

# Putting it all together

- We weigh every $\mathcal{T}_K$ with $W_K$ to indicate its relevance and get the global trust value.
- We get $w_k$ from every given weight $W_K$: $w_k = W_K \cdot rho_K(a, b, c)$, from this we get:

$$\mathcal{T}(a, b, c) = \frac{\sum_{K \in \{I,R,C,W\}} w_K \cdot \mathcal{T}_K(a, b, c)}{\sum_{K \in \{I,R,C,W\}} w_K} \tag{7}$$

- Then the overall reliability becomes:

$$\rho_{\mathcal{T}}(a, b, c) = \frac{\sum_{K \in \{I,R,C,W\}} w_K}{\sum_{K \in \{I,R,C,W\}} W_K} \tag{8}$$

# Summary