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-based 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).

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. This similar to the Regret model.

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) ω_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) ω_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: ω_K
- We use this to express:
  - Rating reliability ρ_{RK}: The total reliability of the individual ratings.
  - Deviation reliability ρ_{DK}: The higher the variability 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.

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

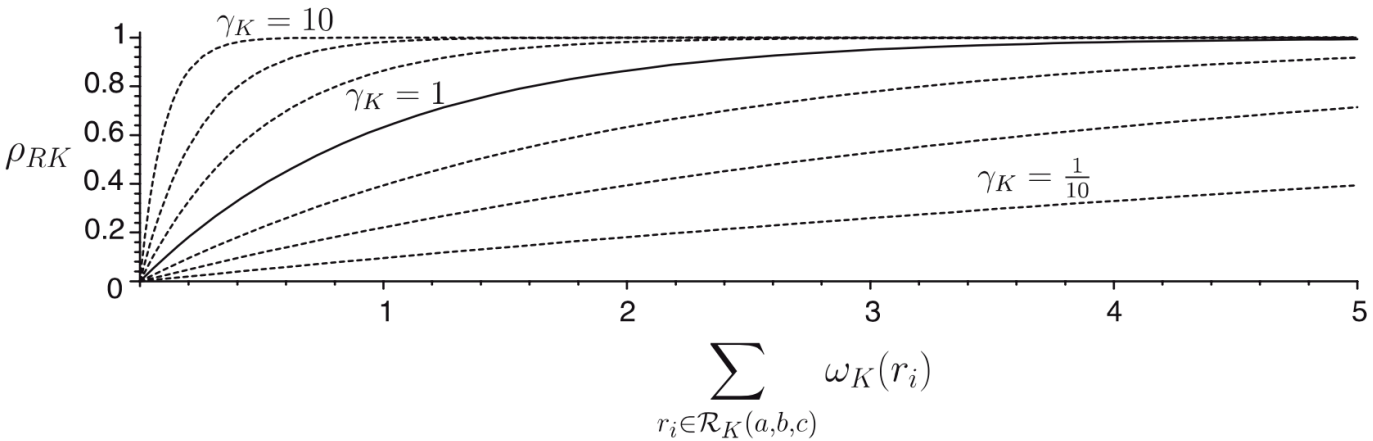

Figure 1: Rating reliability function

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 in the ratings 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 ρ_{RK} and deviation reliability ρ_{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.
- λ is the recency scaling factor.

Interaction trust

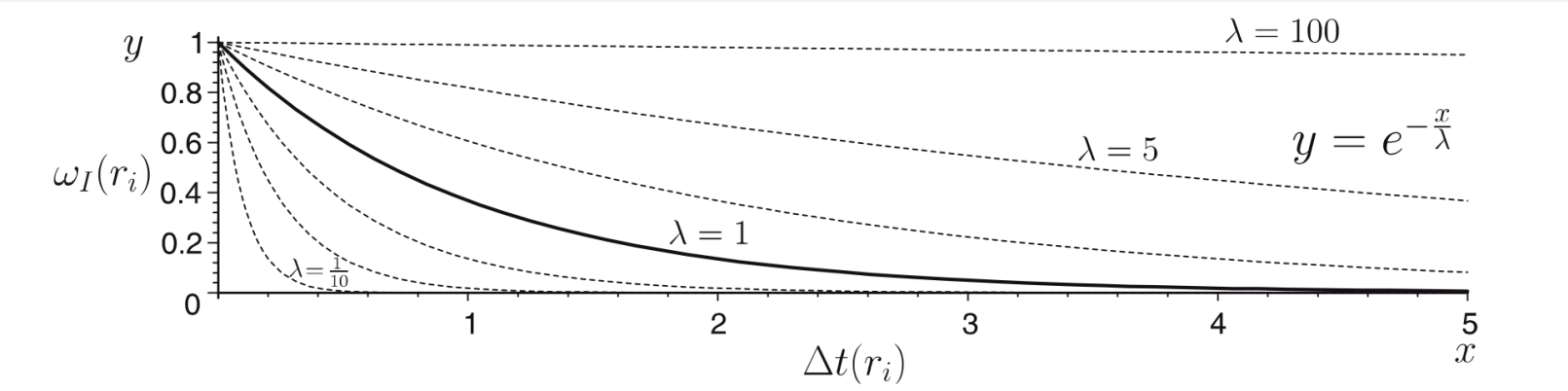

Figure 2: Behavior of the weight function $\omega_I(r_i)$.

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

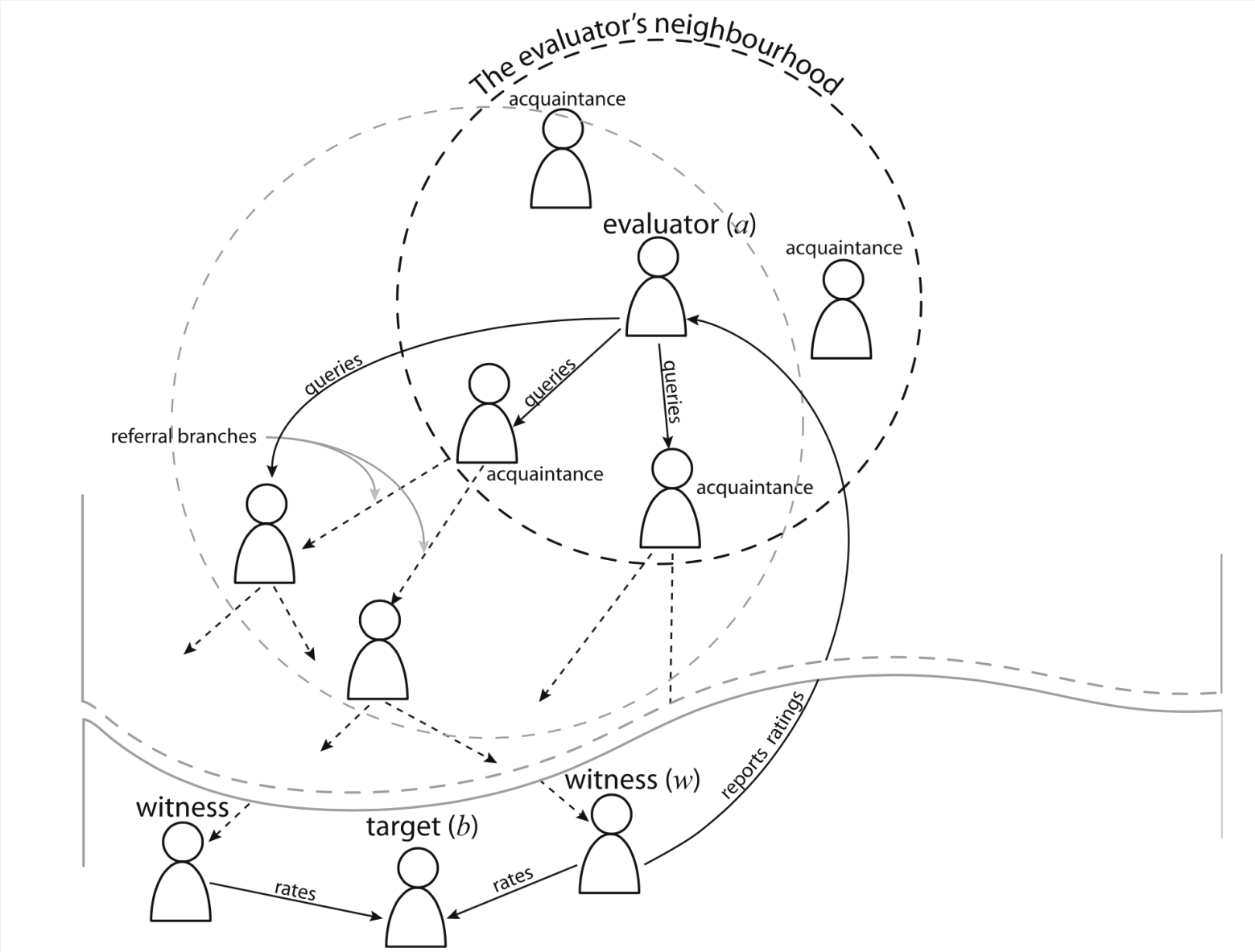

Figure 3: How to find witnesses

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:

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

Now lets test this model…

- Providers: agents which provide a service
  - Four different types of performance: good, ordinary, bad, and intermittent
- Consumers: agents which ask a provider for a service (selection process)
- Act in rounds, not a continuous stream of actions

How to make the environment dynamic:

1. Change the population: add/remove x providers and y consumers randomly
2. Change relationships between agents: change its location in the world
3. Change the behavior of providers: change average performance by a certain amount each round

Questions we want answers on…

1. How does FIRE perform in a static world?
   - Typical situation with 50% good and 50% bad providers
   - Situations with only good or bad providers
2. How does each component of FIRE perform?
3. How does FIRE perform in a dynamic world?

Typical provider population

- Consisting of 50% profitable providers (i.e. yielding positive UG) and 50% exploiting providers (yielding negative UG)
- Static environment

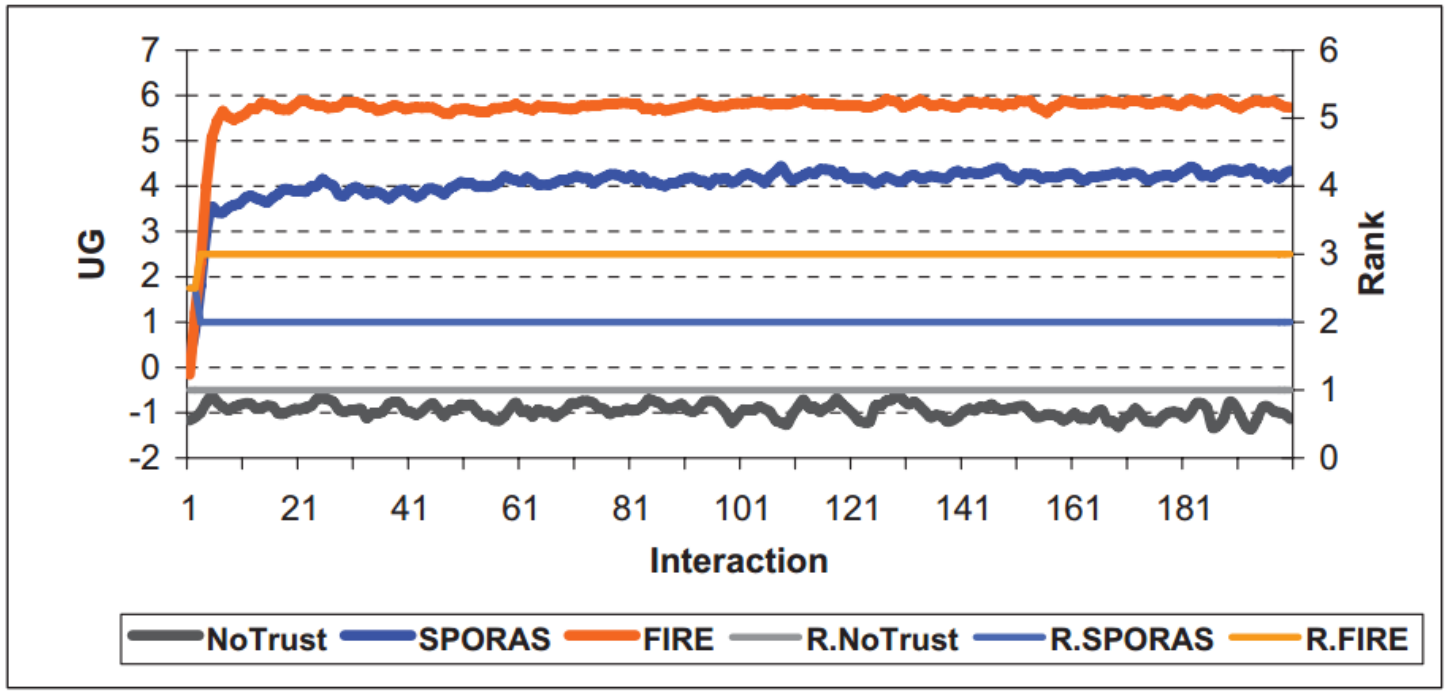

FIGURE 5.1: Overall performance of FIRE in the typical provider population.

Performance of FIRE (100% good providers)

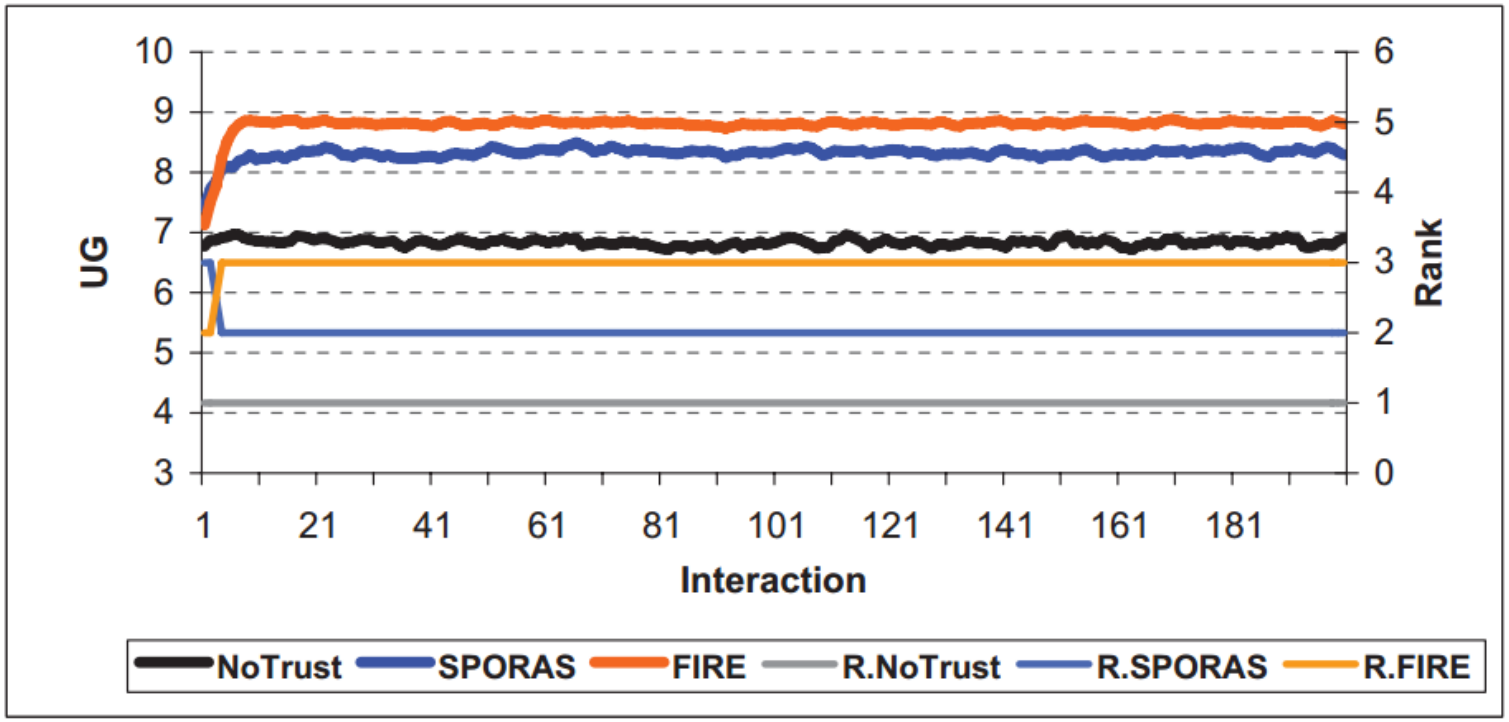

FIGURE 5.2: Overall performance of FIRE – 100% good providers.

Figure 5: Performance good providers

Performance of FIRE (100% ordinary providers)

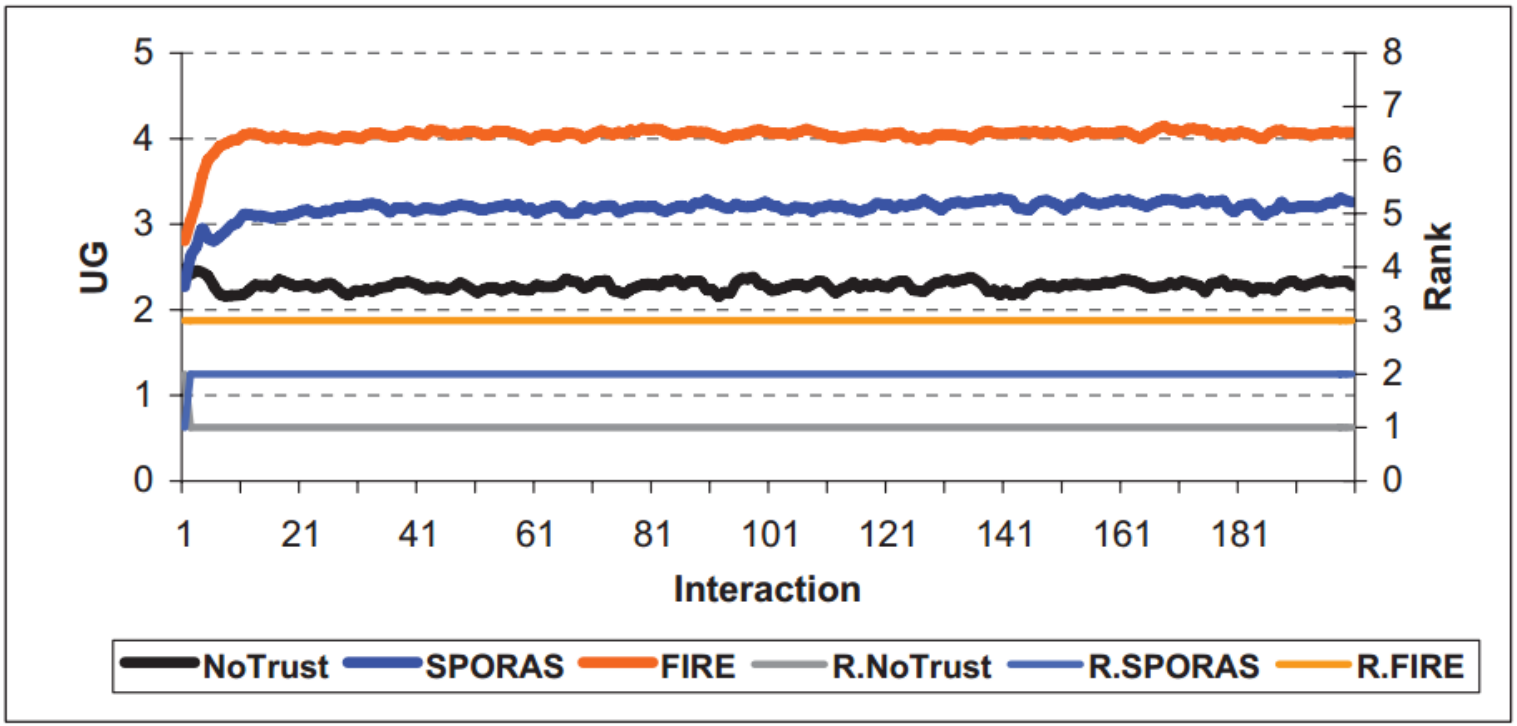

FIGURE 5.3: Overall performance of FIRE – 100% ordinary providers.

Figure 6: Performance ordinary providers

Performance of FIRE (100% bad providers)

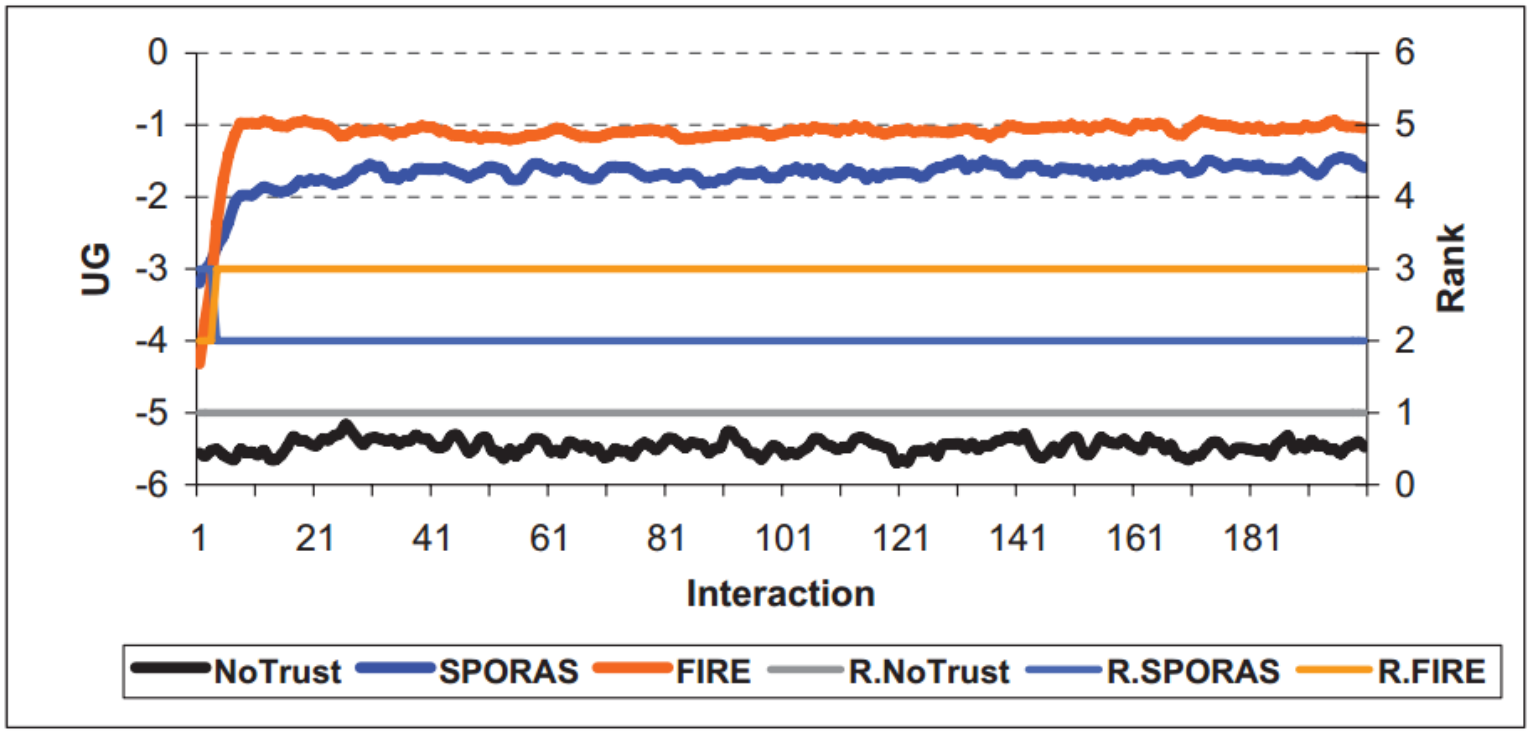

FIGURE 5.4: Overall performance of FIRE – 100% bad providers.

Figure 7: Performance bad providers

Performance of FIRE's novel components (WR)

Since the IT components are mostly reused from Regret, we only look at the novel components; WR and CR.

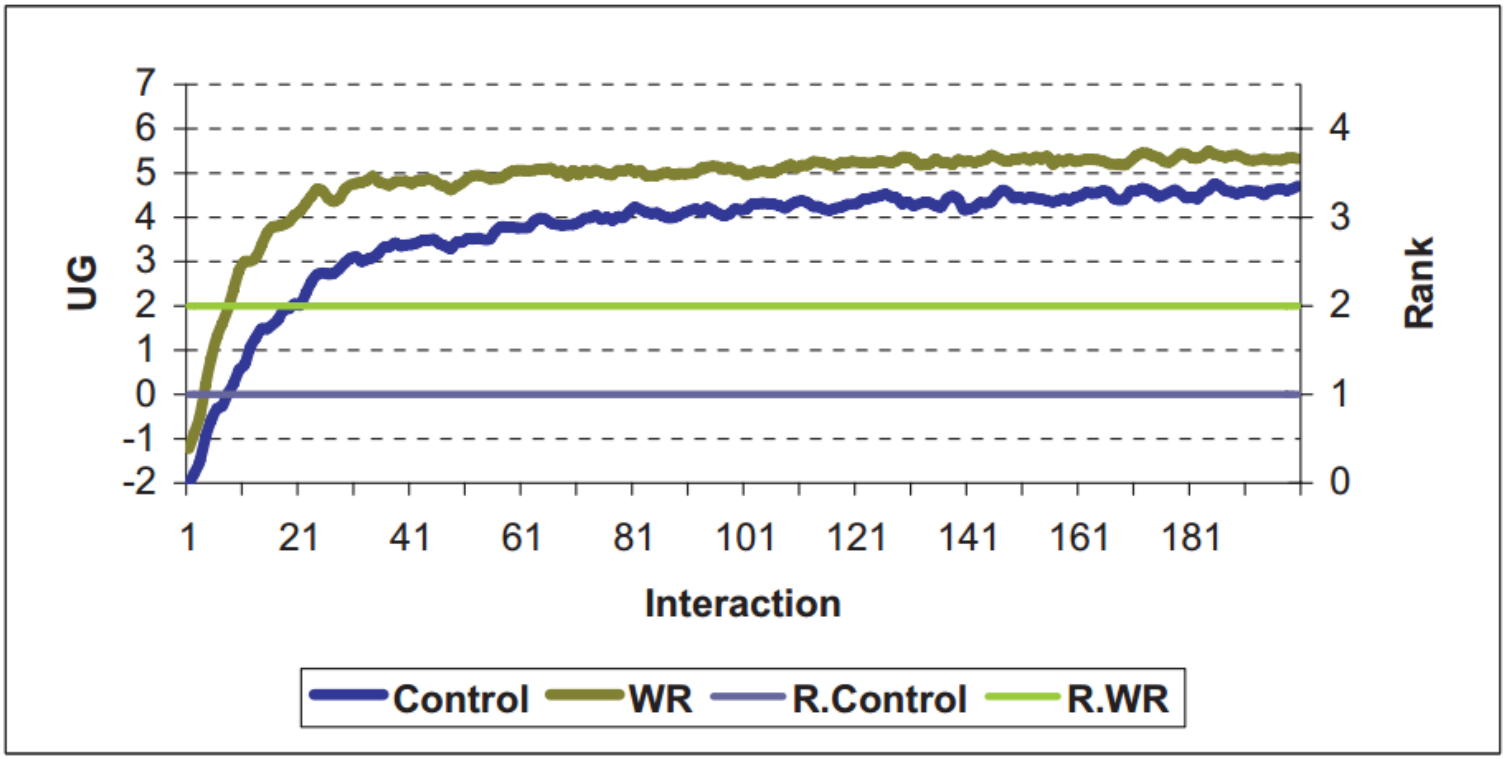

FIGURE 5.12: Performance of the WR component.

Figure 8: Performance WR

Performance of FIRE's novel components (CR)

Since the IT components are mostly reused from Regret, we only look at the novel components; WR and CR.

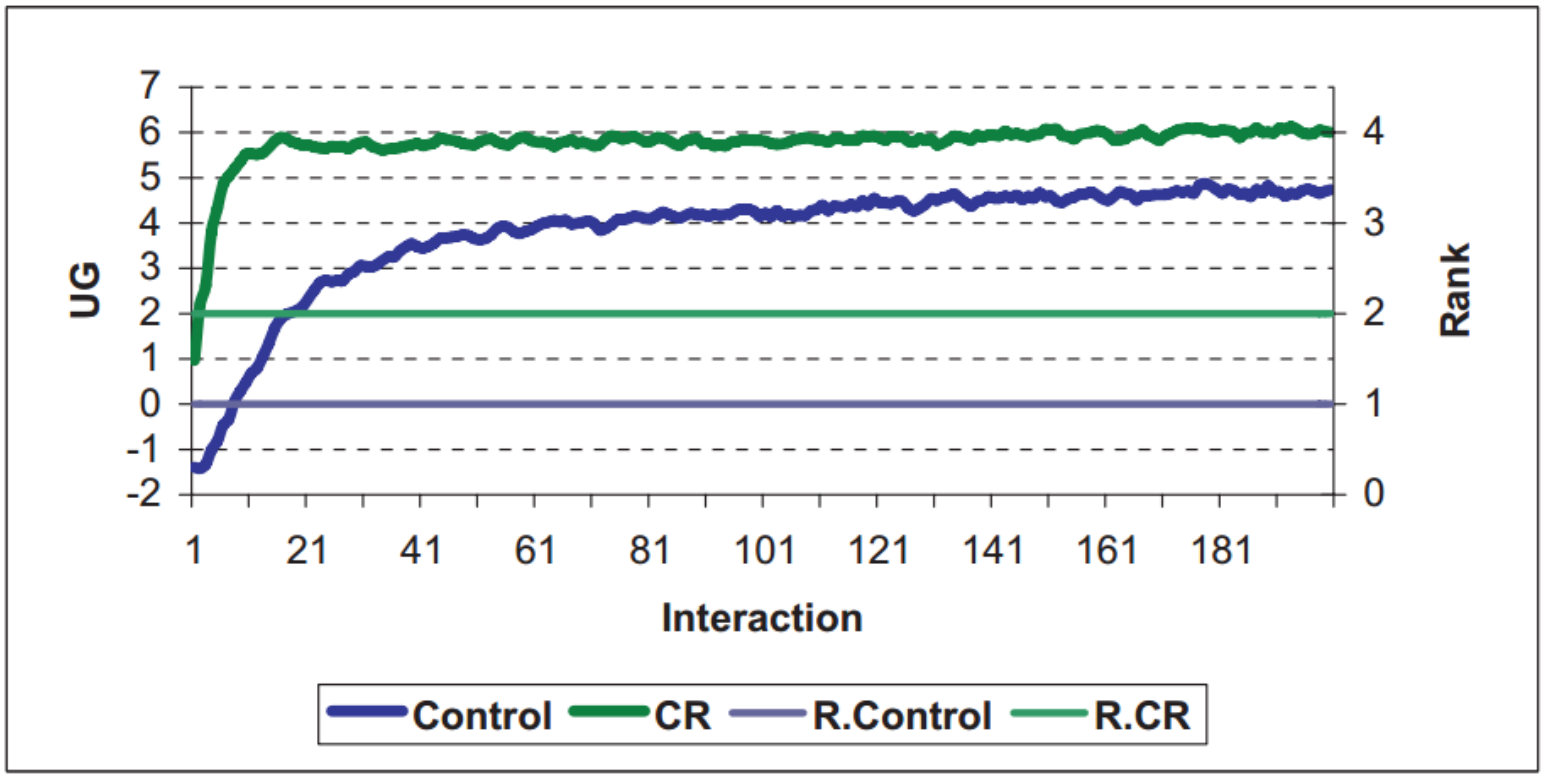

FIGURE 5.13: Performance of the CR component.

Figure 9: Performance CR

Performance of FIRE in a dynamic environment

Several conditions tested, such as…

- The provider population changes at maximum 2% every round
- The consumer population changes at maximum 5% every round
- A provider may switch into a different (performance) profile with a probability of 2% every round And more…

Performance of FIRE in a dynamic environment

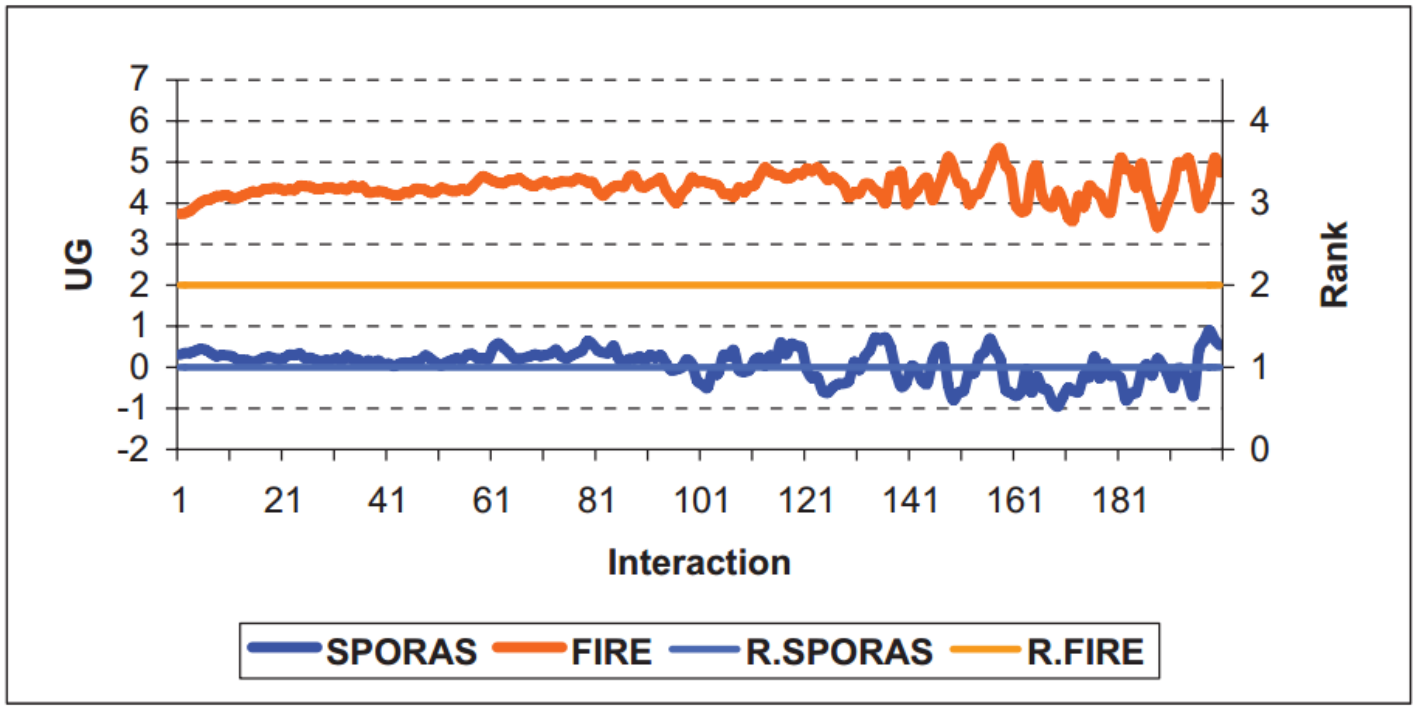

FIGURE 5.11: Experiment 7: Performance of FIRE in an environment where all dynamic factors are in effect.

Figure 10: Performance Dynamic World

What have we seen?

- FIRE introduces a generic framework which combines multiple sources of trust information to provide a collective and precise trust measure.
- Using FIRE, agents are in general better in selecting the best partner, resulting in a better UG.
- FIRE can handle various types of changes in an open MAS very well.
- Specifically, the WR and CR components contribute highly to FIRE's performance.

Are we done?

> “Agents are honest in exchanging information with one another.”

Isn’t this in contradiction with what we want to achieve here?

Without this assumption…

Third-party information can of course be innacurate:

1. One person can see ‘on-time good delivery’ as an excellent service, but someone else can see this as ‘satisfactory’.
2. You can deliberately provide false information about someone, to serve your own interests.

To fix this, they have extended the model :)

The Credibility Model

- Computes the credibility of a witness or a referee, based on the IT components in FIRE
- These measures are called the witness credibility and referee credibility
- The procedures of computing these measures are (almost) the same, I´ll show you witness credibility

Witness Credibility

After having an interaction of agent a with b…

1. a records its rating about b's performance: r_a = (a, b, i_a, c, v_a)
2. When a previously received a witness rating from w about b: r_k = (w, b, i_k, c, v_k)
3. …it rates the credibility v_w of w:

$$v_w = \begin{cases} 1 - |v_k - v_a| & \text{if } |v_k - v_a| < l \\ -1 & \text{if } |v_k - v_a| \geq l \end{cases} \tag{9}$$

How to compute the witness credibility?

$$\mathcal{T}_{WCr}(a,w) = \begin{cases} \mathcal{T}_I(a,w,term_{WCr}) & \text{if } \mathcal{R}_I(a,w,term_{WCr}) \neq \emptyset \\ \mathcal{T}_{WCr} & otherwise \end{cases} \tag{10}$$

$$\omega_W(r_i) = \begin{cases} 0 & \text{if } \mathcal{T}_{WCr}(a,w) \leq 0 \\ \mathcal{T}_{WCr}(a,w) \cdot \omega_I(r_i) & otherwise \end{cases} \tag{11}$$

Computing a referee's credibility has the same approach.

Testing witness inaccuracy level

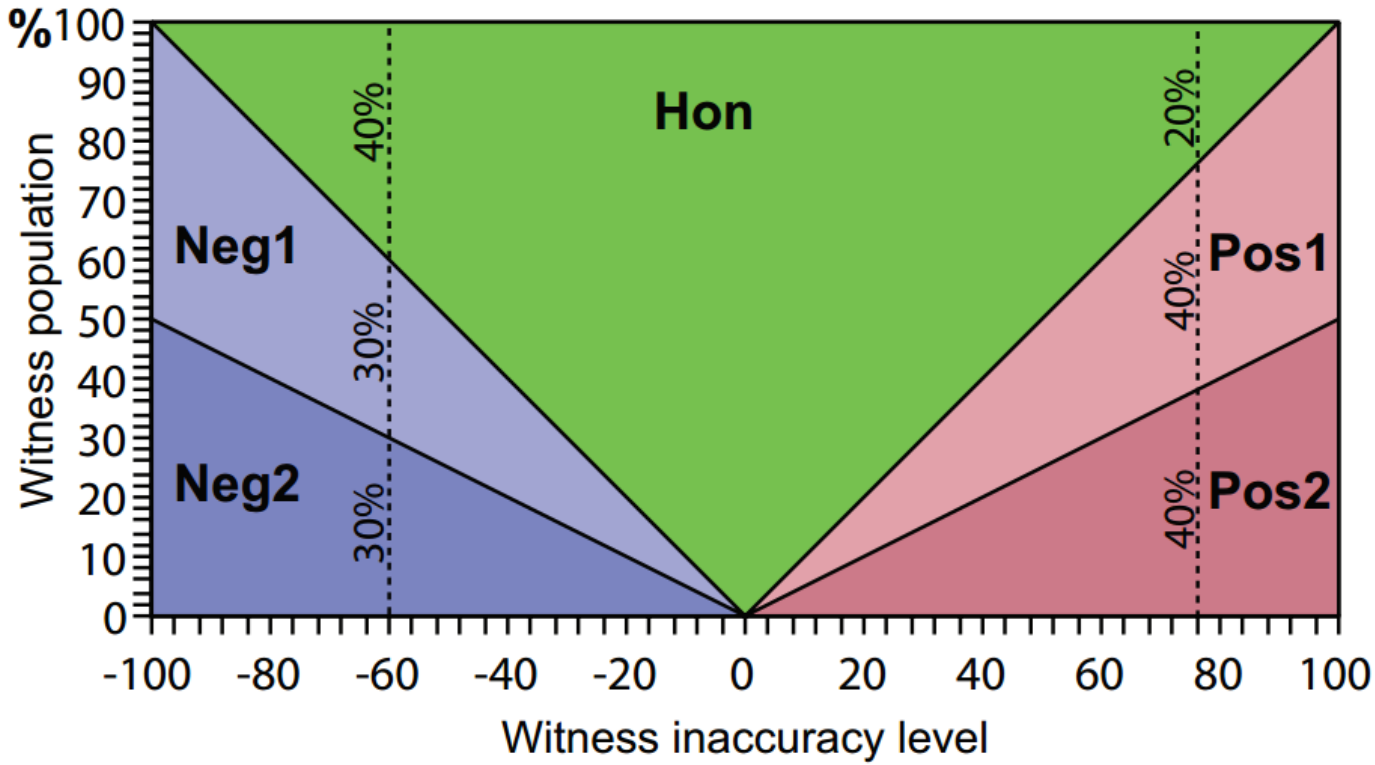

FIGURE 6.1: The proportions of witness types at various levels of witness inaccuracy.

Figure 11: Witness inaccuracy level

Testing witness inaccuracy level

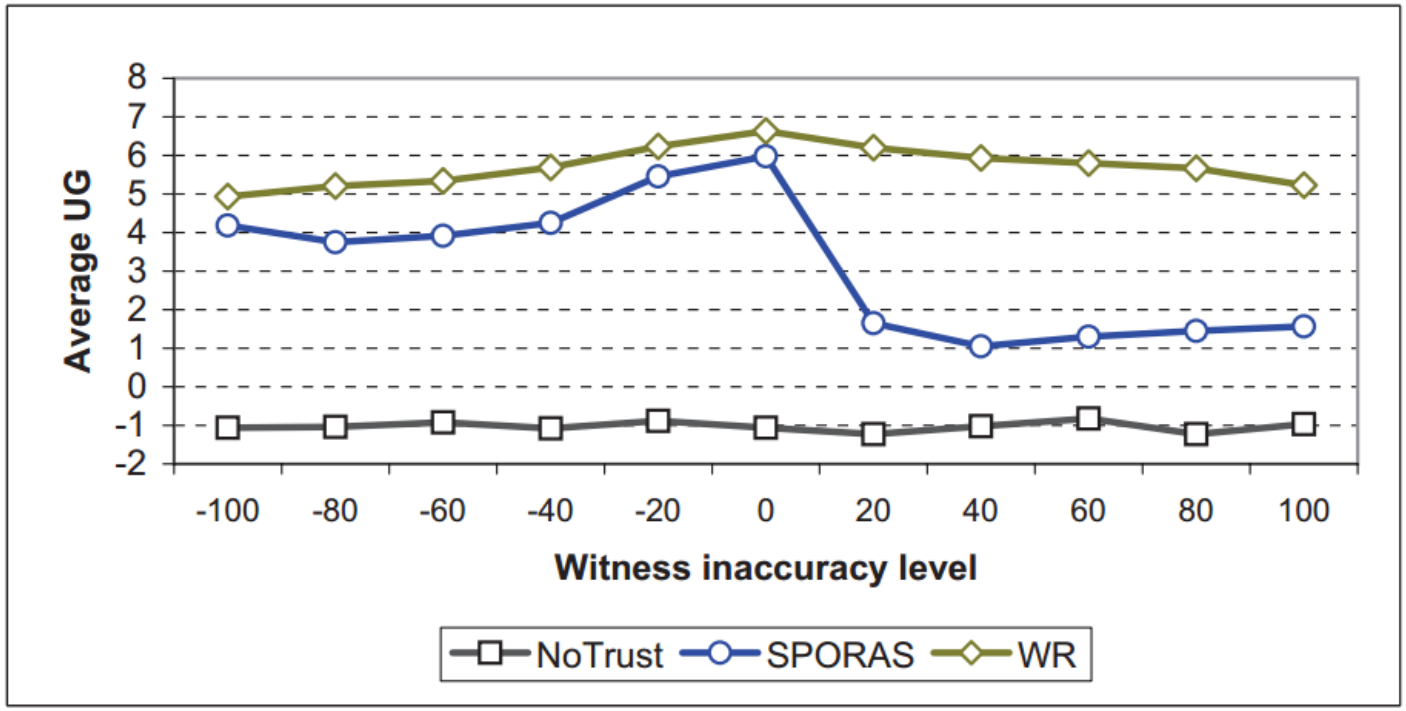

FIGURE 6.3: Performance of NoTrust, SPORAS, and WR at various levels of inaccuracy.

Figure 12: Witness inaccuracy level

Discussion

- We have seen how the different kinds of trust and reputation are combined.
- It shows significant improvement over older systems.
- The WR and CR components contribute highly to FIRE's performance.

Discussion

- We have seen how the different kinds of trust and reputation are combined.
- It shows significant improvement over older systems.
- The WR and CR components contribute highly to FIRE's performance.

- Assumption about truthful communication does not hold in reality.
- How would we deal with malicious agents?
- What component could be added to the model?