Agent-Based Systems:

An Introduction to Agents and an Overview of Recent/Current Research

Nathan Griffiths (nathan@dcs.warwick.ac.uk)

Sarah Lim Choi Keung (slck@dcs.warwick.ac.uk)

Department of Computer Science, University of Warwick, CV4 7AL, UK

Overview of talk

- Part 1: Agent-based systems — a primer
- Part 2: Combining Trust, Reputation and Relationships for Improved Agent Interactions — Sarah’s research
- Part 3: Trust, reputation and clans — a brief overview of my recent work.

Part 1

Agent-based systems

(A brief introduction)

Agents: computing as interaction

Various metaphors have been applied to computing:

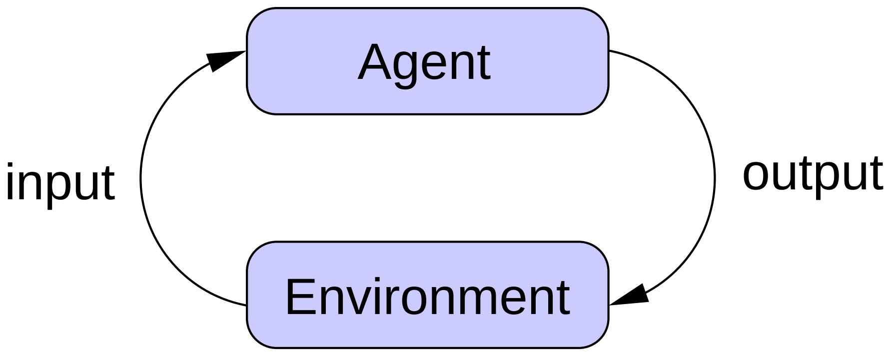
- computation as *calculation*: mainly pre-1960s
- computation as *information processing*: 1960s–present
- computation as *interaction*: 1990s–future

Computation as interaction is an inherently social view, with applications being built from societies of components that may be distributed in geography and ownership. These components are *agents*. . .

[Adapted from Agent Technology: Computing as Interaction, A Roadmap for Agent Based Computing, AgentLink, 2005. *Ask me for a copy!*]

So, what *is* an agent?

An *agent* is an entity which perceives its environment and is able to act, typically autonomously and pro-actively, in order to solve particular problems, whilst remaining responsive to its environment. Agents typically have the ability the interact with other agents, and form cooperating ‘societies’, or *multi-agent systems*.



The rise of agents

Agents have become increasing popular since the 1990s, and have been applied in areas as diverse as:

- Distributed processing and problem solving, e.g. Grid Computing, GeneWeaver, AgentCities.net
- Data mining, e.g. IBM Intelligent Miner
- Critical system monitoring, e.g. Power distribution, air traffic control
- Robotics, e.g. NASA Mars rover
- Entertainment, e.g. Film (LoTR etc.) and computer games (Quake, Counterstrike etc.)

Perspectives on agent technology

Can view agent technology from three perspectives:

- Agents as a *design metaphor*: view applications as a set of autonomous interacting agents; the design problem is to define the entities, communication mechanisms, social norms etc.

Perspectives on agent technology

Can view agent technology from three perspectives:

- Agents as a *design metaphor*: view applications as a set of autonomous interacting agents; the design problem is to define the entities, communication mechanisms, social norms etc.
- Agents as *technologies*: techniques and tools such as architectures that balance reactivity and deliberation, methods for learning and knowledge representation, and communication and negotiation mechanisms.

Perspectives on agent technology

Can view agent technology from three perspectives:

- Agents as a *design metaphor*: view applications as a set of autonomous interacting agents; the design problem is to define the entities, communication mechanisms, social norms etc.
- Agents as *technologies*: techniques and tools such as architectures that balance reactivity and deliberation, methods for learning and knowledge representation, and communication and negotiation mechanisms.
- Agents as *simulation*: a natural way to simulate and model complex and dynamic environments, e.g. economic systems, societies and biological systems.

Properties of agents

There is no precise definition of agency, but it is widely accepted that agents exhibit:

- autonomy: operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state;

[Wooldridge and Jennings, Intelligent Agents. KER 10(2), 1995]

Properties of agents

There is no precise definition of agency, but it is widely accepted that agents exhibit:

- autonomy: operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state;
- social ability: interact with others (and possibly humans) usually via some ACL;

[Wooldridge and Jennings, Intelligent Agents. KER 10(2), 1995]

Properties of agents

There is no precise definition of agency, but it is widely accepted that agents exhibit:

- autonomy: operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state;
- social ability: interact with others (and possibly humans) usually via some ACL;
- reactivity: perceive their environment, and respond in a timely fashion to changes that occur in it;

[Wooldridge and Jennings, Intelligent Agents. KER 10(2), 1995]

Properties of agents

There is no precise definition of agency, but it is widely accepted that agents exhibit:

- autonomy: operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state;
- social ability: interact with others (and possibly humans) usually via some ACL;
- reactivity: perceive their environment, and respond in a timely fashion to changes that occur in it;
- pro-activeness: able to exhibit goal-directed behaviour by taking the initiative.

[Wooldridge and Jennings, Intelligent Agents. KER 10(2), 1995]

Other possible properties

Some argue that agents should also exhibit:

- mobility: the ability to move around their (possibly software) environment;
- veracity: will not knowingly communicate false information;
- benevolence: agents do not have conflicting goals, and each tries to do what is asked of it;
- rationality: act in order to achieve its goals, and will not act in such a way as to prevent its goals being achieved.

These properties are contentious and are NOT universally accepted.

Our view

We take the view that agents are autonomous, social, reactive, pro-active and are *self-interested*, i.e. they are:

- *not* veracious: agents will not necessarily tell the truth;
- *not* benevolent: will not necessarily do what is asked of them, and furthermore may act to maliciously prevent another from achieving its goals.

Additionally, agents may be heterogeneous.

Multi-Agent Systems (MAS)

- Collection of interacting agents:
  - to achieve some collective goal (unachievable individually) through cooperation, or
  - agent situated in a multi-agent environment where achievement of its goals involves interaction with others, e.g. automated trading

Multi-Agent Systems (MAS)

- Collection of interacting agents:
  - to achieve some collective goal (unachievable individually) through cooperation, or
  - agent situated in a multi-agent environment where achievement of its goals involves interaction with others, e.g. automated trading
  - closely coupled vs. loosely coupled viewpoint
  - may be cooperative or competitive

Multi-Agent Systems (MAS)

- Collection of interacting agents:
  - to achieve some collective goal (unachievable individually) through cooperation, or
  - agent situated in a multi-agent environment where achievement of its goals involves interaction with others, e.g. automated trading
  - closely coupled vs. loosely coupled viewpoint
  - may be cooperative or competitive
- Issues include diverse goals, capabilities, and beliefs, establishing cooperation, managing risk
- Agents must cooperate and coordinate efforts to be successful.

Part 2

Combining Trust, Reputation and Relationships for Improved Agent Interactions

(An overview of Sarah's work)

Part 3

Trust, reputation and clans

(An overview of my recent work)

Context of the research

- Distributed systems generally rely on the interactions of many autonomous agents
- Agents typically have specific individual capabilities, knowledge and resources, and vary in reliability, quality and honesty

Context of the research

- Distributed systems generally rely on the interactions of many autonomous agents
- Agents typically have specific individual capabilities, knowledge and resources, and vary in reliability, quality and honesty
- Autonomy gives rise to *uncertain* interactions, with a risk of task failure, lateness or increased cost
- The problems: how to engender cooperation and delegate tasks to agents appropriately, e.g. minimise cost and risk of failure while maximising quality and timeliness.

Possible solutions

Possible solutions to these problems:

- Multi-dimensional trust (MDT)
- Combining trust and reputation (MDT-R)
- Fuzzy trust
- Clans: medium term coalitions

Basic concept: trust

- *Trust* is an estimate of how likely an agent is to fulfil its commitments, i.e. it is an estimate of *risk*
  - experience-based: result of individual experiences
  - recommendation-based: result of information provided by others

Basic concept: trust

- *Trust* is an estimate of how likely an agent is to fulfil its commitments, i.e. it is an estimate of *risk*
  - experience-based: result of individual experiences
  - recommendation-based: result of information provided by others
- Experience-based trust can be naturally applied to the full range of distributed systems

Basic concept: trust

- *Trust* is an estimate of how likely an agent is to fulfil its commitments, i.e. it is an estimate of *risk*
  - experience-based: result of individual experiences
  - recommendation-based: result of information provided by others
- Experience-based trust can be naturally applied to the full range of distributed systems
- Recommendation-based trust is more powerful, but
  - can be a lack of motivation to offer information (risk that good feedback leads to “swamping”)
  - (also, issues of subjectivity and situational feedback).

Recent research strands

- Multi-dimensional trust (MDT)
  (AAMAS 2005; iTrust 2005)
- Combining trust and reputation (MDT-R)
  (CSCWD 2005; ESWA 31 (4), 2006)
- Fuzzy trust
  (DAKS 2006; CIA 2006)
- Clans: medium term coalitions
  (Kybernetes 34 (9/10), 2005)

MDT: Trust representation

- We can represent trust T in an agent α, to be a real number in the interval between 0 (distrust) and 1 (blind trust): $T_\alpha \in [0, 1]$
- The numbers are merely comparative values, with no strong implied semantics

MDT: Trust representation

- We can represent trust T in an agent α, to be a real number in the interval between 0 (distrust) and 1 (blind trust): $T_\alpha \in [0, 1]$
- The numbers are merely comparative values, with no strong implied semantics
- Initial trust value $T_{initial}$ determined by the agent's disposition: optimistic $\Leftrightarrow$ pessimistic (former ascribes high values; the latter low values)

MDT: Trust representation

- We can represent trust T in an agent α, to be a real number in the interval between 0 (distrust) and 1 (blind trust): $T_\alpha \in [0, 1]$
- The numbers are merely comparative values, with no strong implied semantics
- Initial trust value $T_{initial}$ determined by the agent's disposition: optimistic $\Leftrightarrow$ pessimistic (former ascribes high values; the latter low values)
- Multi-dimensional Trust: decomposed according to dimensions of interaction, e.g. success T_α^s, cost T_α^c, timeliness T_α^t and quality T_α^q, etc.
- $T_\alpha^s, T_\alpha^c, T_\alpha^t, T_\alpha^q \in [0 : 1]$.

MDT: Maintaining trust values

- Trust disposition also defines how trust is updated after interactions — trust increases after successful interactions, and decreases after failure
- Update functions are heuristics, e.g. might use

$$update_{success}(T^d_\alpha) = T^d_\alpha + ((1 - T^d_\alpha) \times (\omega_s \times T^d_\alpha))$$

$$update_{fail}(T^d_\alpha) = T^d_\alpha - ((1 - T^d_\alpha) \times (\omega_f \times T^d_\alpha))$$

  where ω_s and ω_f weights defined by disposition
- Confidence in trust C^d_α increases with experience.

MDT: Trust decay

- Trust values become outdated if experiences no longer relevant: unless reinforced, the positive effect of success reduces over time, as does the negative effect of failure
- Trust values *decay* by converging toward initial value

$$decay_{trust}(T_{\alpha}^{d}) = T_{\alpha}^{d} - \frac{T_{\alpha}^{d} - T_{initial}}{\omega_{td}}$$

- Similarly, confidence in trust values decreases as experience becomes dated

$$decay_{confidence}(C_{\alpha}^{d}) = C_{\alpha}^{d} - \frac{C_{\alpha}^{d}}{\omega_{cd}}$$

- Frequency of decay and weights ω_{td} and ω_{cd} determined by disposition.

MDT: Stratified trust

- An alternative to numerical representation of trust, is to divide the trust continuum into strata, e.g. “very trustworthy”, “trustworthy”, “untrustworthy”, “very untrustworthy” [Abdul-Rahman & Hailes]

MDT: Stratified trust

- An alternative to numerical representation of trust, is to divide the trust continuum into strata, e.g. “very trustworthy”, “trustworthy”, “untrustworthy”, “very untrustworthy” [Abdul-Rahman & Hailes]
  - Key advantage: ease of comparisons — the problem of defining meaning of a numerical value is avoided; avoids overfitting (e.g. distinguishing between trust values of 0.5 and 0.50001); avoids overly narrow set of trusted agents

MDT: Stratified trust

- An alternative to numerical representation of trust, is to divide the trust continuum into strata, e.g. “very trustworthy”, “trustworthy”, “untrustworthy”, “very untrustworthy” [Abdul-Rahman & Hailes]
  - Key advantage: ease of comparisons — the problem of defining meaning of a numerical value is avoided; avoids overfitting (e.g. distinguishing between trust values of 0.5 and 0.50001); avoids overly narrow set of trusted agents
  - Main disadvantages: a loss of sensitivity and accuracy as comparisons become coarse grained (unable to distinguish within a stratum); how to update trust; semantics are still subjective.

MDT: Stratified trust

- Ideal approach combines ease of comparison (stratification) with accuracy/sensitivity (numerical)

MDT: Stratified trust

- Ideal approach combines ease of comparison (stratification) with accuracy/sensitivity (numerical)
- Solution: *variable stratification*: use numerical representation to preserve accuracy and sensitivity and translate trust values into strata immediately before comparison
- Variable number of strata gives the selecting agent flexibility of comparison advantage versus precision.

MDT: Task delegation — combining trust dimensions

- Task Delegation: selecting best agent based on trust and factors such as cost, quality etc.
- These factors are potentially in conflict, e.g. high quality = expensive

MDT: Task delegation — combining trust dimensions

- Task Delegation: selecting best agent based on trust and factors such as cost, quality etc.
- These factors are potentially in conflict, e.g. high quality = expensive
- We use a weighted product model (a standard multi-criteria decision making technique) to combine choice factors (including trust)
- For each agent to which a task could be delegated, calculate performance value by combining factors f_{α_i} $PV(\alpha) = \prod_{i=1}^{n} (f_{\alpha_i})^{\mu_i}$ where μ_i's are weights that sum to 1
- The best delegate is the one with the highest PV.

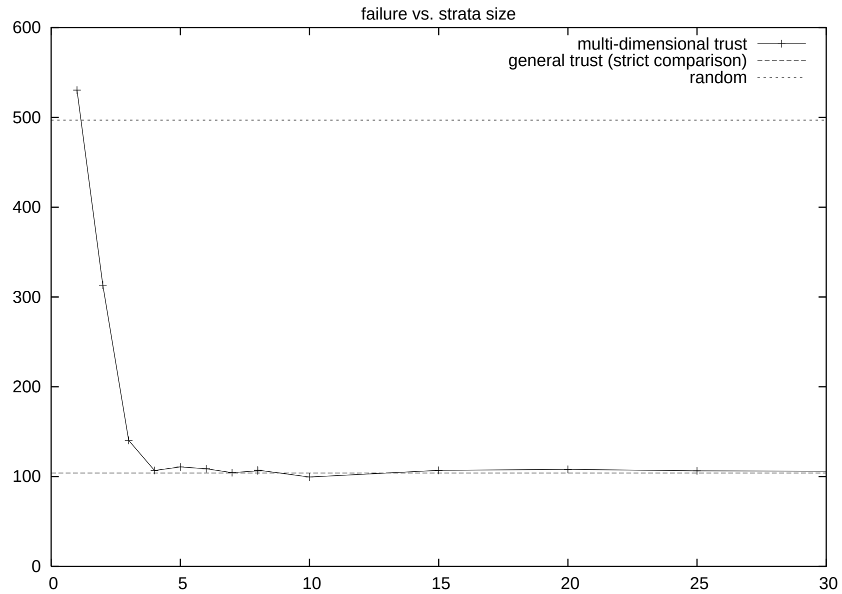
MDT: Task delegation — combining trust dimensions

- Weights give an agent flexibility to delegate tasks according to its current preferences: no change of underlying trust model/data needed

MDT: Task delegation — combining trust dimensions

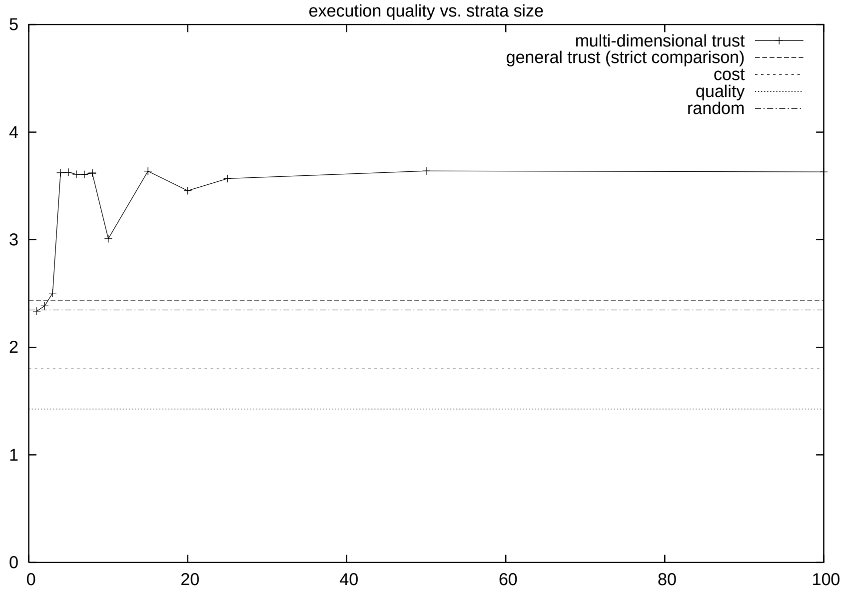
- Weights give an agent flexibility to delegate tasks according to its current preferences: no change of underlying trust model/data needed
- Trust is stratified *before* inclusion:
  $stratify(t) = \lceil t \times s \rceil$
  where s is number of strata chosen by delegating agent
- Factors that should be minimised, such as cost, can be included by using
  $f_{\alpha_c} = max(\alpha_c \ldots \xi_c) + 1 - \alpha_c$

MDT: Example results — failure rate in Grid environment



- Equal weighting to factors, slight emphasis on success
- For >10 strata, MDT and strict trust give joint lowest failure rate
- For 2-10 strata, MDT improves from $\approx$ random to strict trust, using wider set of agents

MDT: Example results — execution quality



- For >3 strata MDT gives highest quality, by ≈ 30%
- Quality-based approach is worst: it is based on *advertised* quality not actual or *expected* quality
- Unreliable agents: in these experiments high advertised quality agents tended to be unreliable, and yield lower than advertised quality.

Recent research strands

- Multi-dimensional trust (MDT)
  (AAMAS 2005; iTrust 2005)
- Combining trust and reputation (MDT-R)
  (CSCWD 2005; ESWA, 31 (4), 2006)
- Fuzzy trust
  (DAKS 2006; CIA 2006)
- Clans: medium term coalitions
  (Kybernetes 34 (9/10), 2005)

MDT-R: Trust representation

- MDT-R: extend MDT by combining trust with *recommendations*, i.e. experience and recommendation based trust
- Again, trust in each dimension a real number in $[0:1]$
- Update and decay functions (relatively) unchanged.

MDT-R: Interaction summaries

- To share trust information, need a clear semantics
- Existing approaches tend to suffer from subjectivity

MDT-R: Interaction summaries

- To share trust information, need a clear semantics
- Existing approaches tend to suffer from subjectivity
- Solution: *communicate interaction summaries* not trust values: agent α reveals number of interactions with γ in which expectations met $I_{\alpha\gamma}^{d+}$ and number in which not met $I_{\alpha\gamma}^{d-}$ in each dimension d

MDT-R: Interaction summaries

- To share trust information, need a clear semantics
- Existing approaches tend to suffer from subjectivity
- Solution: *communicate interaction summaries* not trust values: agent α reveals number of interactions with γ in which expectations met $I_{\alpha\gamma}^{d+}$ and number in which not met $I_{\alpha\gamma}^{d-}$ in each dimension d
- When delegating trust, agent asks all trusted peers for recommendations (i.e. $I_{\alpha\gamma}^{d+}$ and $I_{\alpha\gamma}^{d-}$)
- Recommendation obtained by summing proportion of interactions where expectations met, weighted by extent of experience.

$$R_{\gamma}^{d} = \sum_{i=\alpha}^{\xi} \left(\frac{I_{i\gamma}^{d+}}{I_{i\gamma}^{d+} + I_{i\gamma}^{d-}} \times \frac{I_{i\gamma}^{d+} + I_{i\gamma}^{d-}}{total\_interactions} \right)$$

MDT-R: Task delegation — combining trust dimensions

- Again, use a weighted product model to combine choice factors, including trust and recommendations
- Experimental results: MDT-R provides improvement of up to a 30% in achieved quality, and up to a 20% decrease in failure rate, over “traditional” delegation methods using advertised cost and quality of peers.

Recent research strands

- Multi-dimensional trust (MDT)
  (AAMAS 2005; iTrust 2005)
- Combining trust and reputation (MDT-R)
  (CSCWD 2005; ESWA, 31 (4), 2006)
- Fuzzy trust
  (DAKS 2006; CIA 2006)
- Clans: medium term coalitions
  (Kybernetes 34 (9/10), 2005)

Fuzzy Trust: An alternative representation

- Although based on known outcomes of experiences, trust is inherently uncertain: fuzzy logic offers the ability to handle uncertainty and imprecision
- Membership of a classical set is clearly defined, e.g. a person of age 10 might be a member of the set $young$, and not of the set old
- However, the *concept* of young is imprecise
- *Fuzzy sets* have a *membership function*, $\mu(x)$, which defines the degree of membership ($[0:1]$)
- For example, age 35 might have a membership of 0.8 in a fuzzy set $\widetilde{y}$, representing young ages, and a 0.1 membership in the set $\widetilde{o}$ representing old ages.

Fuzzy Trust: Trust representation

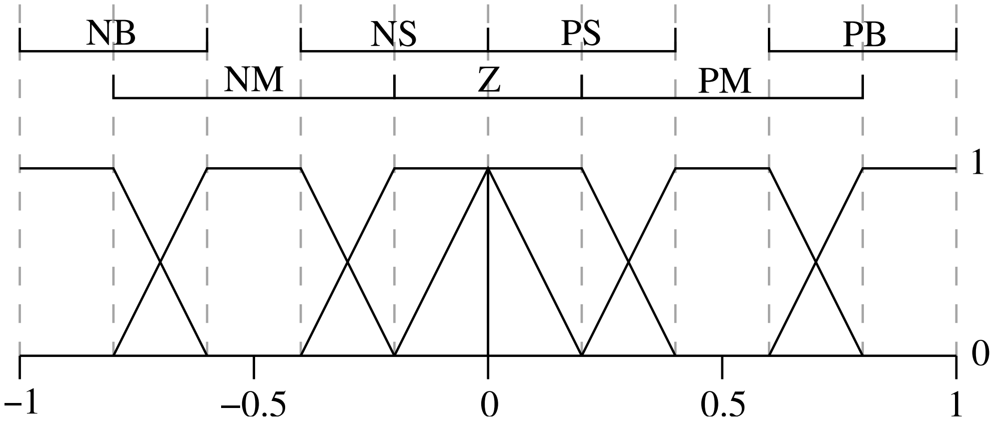
- Again, take a multi-dimensional approach where agents maintain a history of their interactions
- The *experience*, e_α^d, in each dimension d, for each agent α, can be calculated as:

$$e_\alpha^d = \frac{I_\alpha^{d+} - I_\alpha^{d-}}{I_\alpha^{d+} + I_\alpha^{d-}}$$

- Old interactions are purged to ensure relevance
- These experiences are crisp values that must be fuzzyfied to reason about trust.

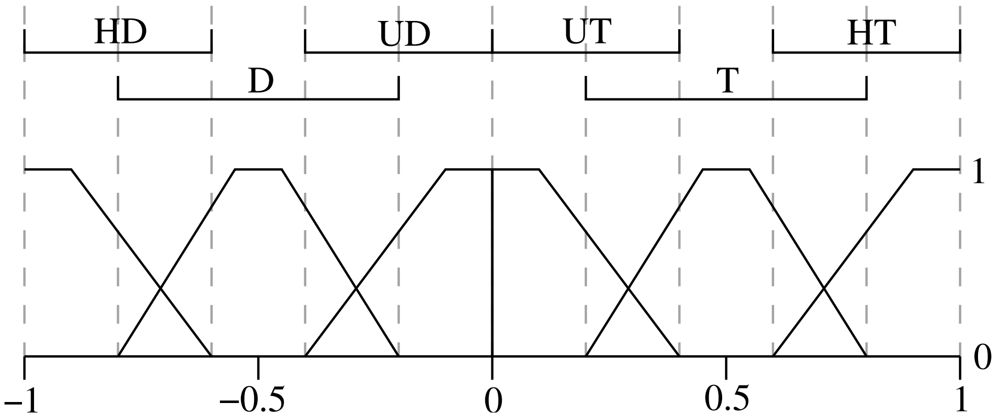
Fuzzy Trust: Trust representation

We define fuzzy terms for experience:



[N=negative, Z=zero, P=positive, B=big, M=medium, S=small]

and trust:

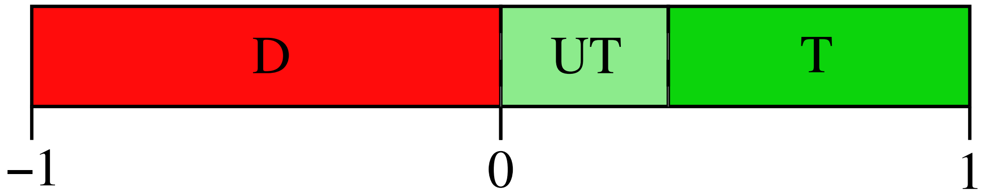


[D=distrust, T=trust, H=high, U=un-]

Fuzzy Trust: A richer view trust

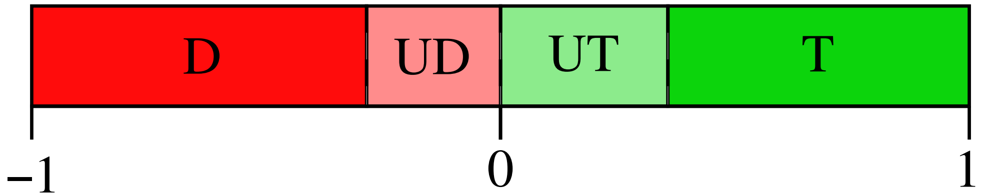
Introduce new notions of:

untrust [Marsh]



and

undistrust [Griffiths]



Fuzzy Trust: Fuzzy rules for reasoning

We define a set of fuzzy inference rules, e.g.:

($R_{UT}1$) **if** $confidence^d_\alpha < \widetilde{minConfidence}$ **and** E^d_α is $positive$
then T_α is $untrust$

...

(R_T1) **if** E^d_α is $\widetilde{negativeBig}$ **then** T_α is $\widetilde{highDistrust}$

...

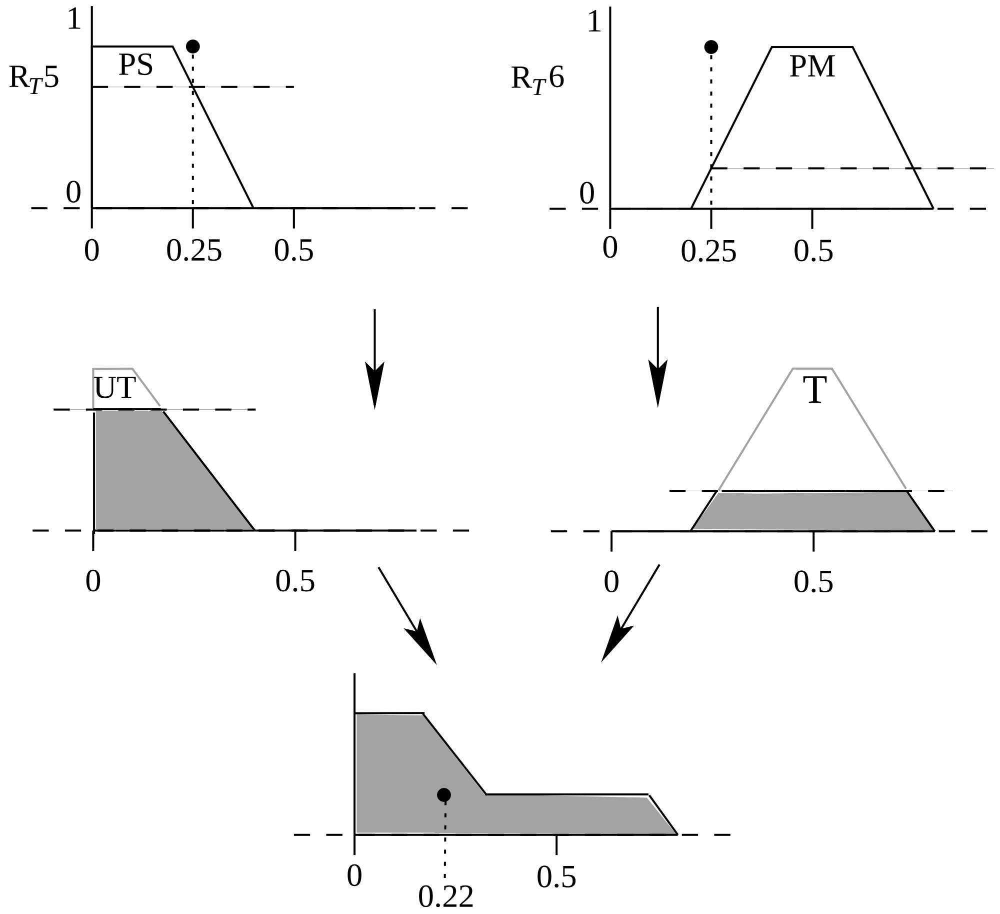
(R_T5) **if** E^d_α is $\widetilde{positiveSmall}$ **then** T_α is $\widetilde{untrust}$

(R_T6) **if** E^d_α is $\widetilde{positiveMedium}$
then T_α is very $\widetilde{trust}$ or $\widetilde{untrust}$

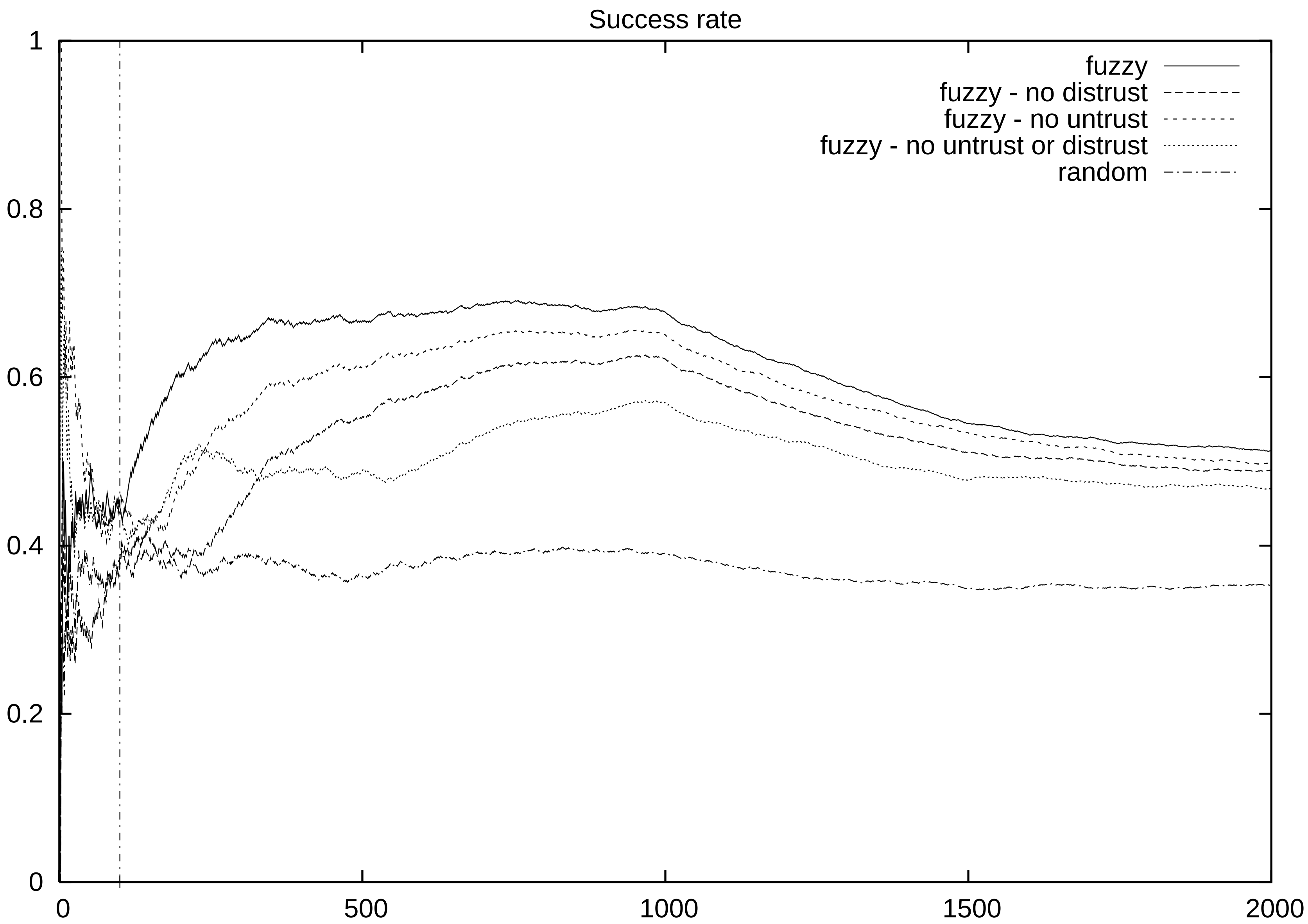
...

Fuzzy Trust: Fuzzy inference

Use Mamdani min-max inference: clip membership degree of conclusions based on membership values of the intersections of the antecedents.



Fuzzy Trust: Example results — success rate



Recent research strands

- Multi-dimensional trust (MDT)
  (AAMAS 2005; iTrust 2005)
- Combining trust and reputation (MDT-R)
  (CSCWD 2005; ESWA, 31 (4), 2006)
- Fuzzy trust
  (DAKS 2006; CIA 2006)
- Clans: medium term coalitions
  (Kybernetes 34 (9/10), 2005)

Clans: Overview

Provide a mechanism for self-interested agents to form clans (coalitions) with trusted agents having similar objectives.

Limitations of previous approaches:

- Agents need common goal *at time of formation*, even if cooperation would be beneficial long-term
- Must recreate group for subsequent tasks
- Problems relating to scalability
- General limitations: *no consideration of trust or motivation*.

Clans: When to form?

- If missing opportunities for cooperation:
  - outgoing requests made by the agent declined
  - incoming requests declined for motivational reasons
- If scalability is a problem (too many agents to consider)
- Lack of information
  - insuffi cient information on others' trustworthiness or capabilities
- Experiencing high failure rates.

Clans: Formation

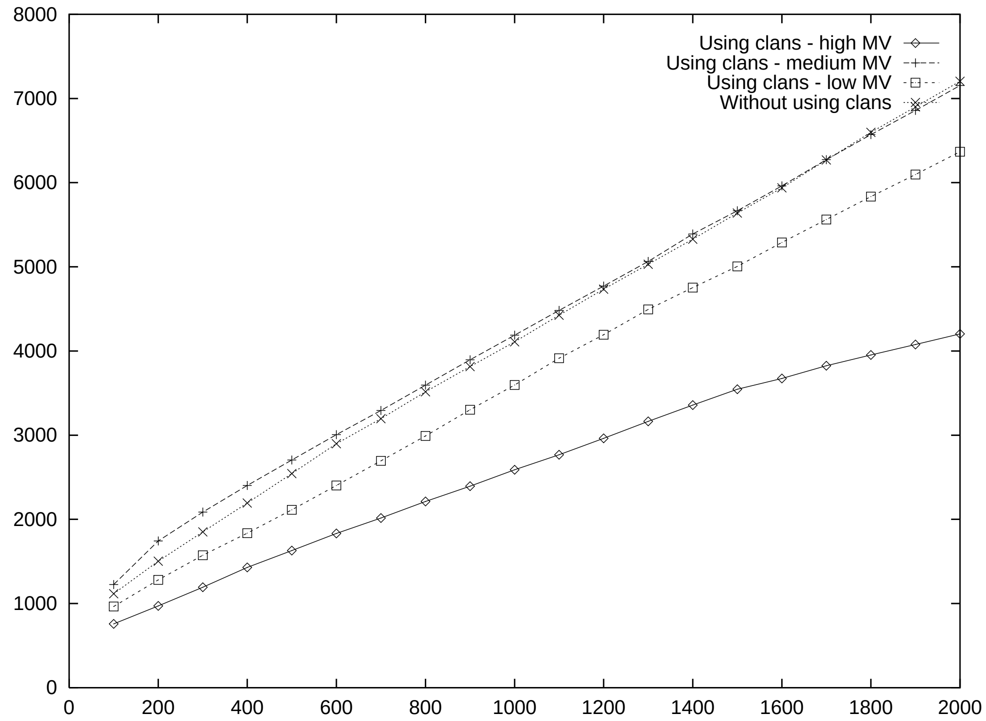
- Requester
  - use current motivations to predict future intentions
  - filter known agents according to their capabilities
  - send request to n most trusted agents, where n is average number of agents needed (scaled for redundancy)
  - as “incentive” include characterisation of current activities

Clans: Formation

- Requester
  - use current motivations to predict future intentions
  - filter known agents according to their capabilities
  - send request to n most trusted agents, where n is average number of agents needed (scaled for redundancy)
  - as “incentive” include characterisation of current activities
- Requestee
  - check requester is trusted
  - consider missed opportunities, scalability, lack of information, and high failure rate —value of joining clan *in general*
  - consider the goals contained in request —value of joining *this specific clan*.

Clans: Reasoning

- Introduce a *kinship motivation*, mitigated by assisting other clan members
- Gives a reduction in failure rate, provided kinship given sufficient importance.



Summary

- Introduced agents (as seen by the MAS community)
- Introduced some key topics:
  - Trust
  - Reputation
  - Relationships
  - Coalitions
- Brief overview of current and recent work.