Engineering Trust in Complex System through Mediating Infrastructures

Alessandro Ricci, *Andrea Omicini*

(aricci@deis.unibo.it, andrea.omicini@unibo.it)

DEIS, Sede di Cesena

Alma Mater Studiorum, Università di Bologna

Outline

- Trust in agent society: state of the art
  - Individual-level and Systemic-level
- Engineering Trust: some missing points
  - Social Trust, Trust in societies, Constructive Trust, Trust & Organisation
- Infrastructure support for Trust
  - mediating infrastructures for Trust creation and management
- Embedding Trust in first class infrastructural abstractions
  - Coordination artifacts
  - Agent Coordination Contexts

Complex Systems & Trust

- An issue in complex systems
  - Users point of view
    - Critical systems (e-Commerce, e-Government, e-Healthcare,...)
  - Engineers point of view
    - Governing complexity
- *Agent societies* as engineering paradigm for complex systems
  - openness, heterogeneity, dynamism
- Trust outside & inside agent societies
  - users and engineers vs. society
  - agents vs. society
  - society vs. agents

Our objectives

- Trust & complex system engineering
  - beyond classic trading scenarios
  - beyond reductionistic conception of Trust
- Infrastructure support for Trust
  - trust engineering
- Trust in (agent) Institutions and infrastructures

Trust in MAS

- *"A belief an agent has that the other party will do what it says it will (being honest and reliable) or reciprocate (being reciprocative for the common good or both), given the opportunity to defect to get higher payoff"*
  - Trust evaluation & revision
- Two main categories (*)
  - Individual-level Trust
    - all the burden on agents
  - System-level trust
    - system *rules of encounter* forcing agents to be trustworthy

(*) Ramchurn, Hunyh, Jennings - "Trust in MAS" - Knowledge Engineering Review, 2004

Individual-level Trust

- *Models* adopted by agents to choose reliable interaction partners
  - open environments
  - deliberating strategies
- Three main models
  - Learning (evolution) based
  - Reputation based
  - Socio-cognitive based
- Problems
  - computationally expensive
  - gathering information in open environments

System-level Trust

- *Mechanisms* dictating rules of encounter to force agents to be trustworthy
  - auctions, voting, market-mechanisms,...
- Three main mechanisms
  - trustworthy interaction mechanisms
  - reputation mechanisms
  - distributed security mechanisms
- Individual & System-level synergy
  - models for enabling agent reasoning
  - mechanisms to ensure that actions can be trusted

Missing points: the engineering perspective

- Wider perspective
  - humans vs. systems
    - users vs. MASs
    - designers / engineers vs. MAS
  - systems vs. systems
    - individual agents vs. society
- Levels
  - Social trust
  - Trust in societies
  - Constructive trust
  - Trust and organisation

Mediating Infrastructures for Trust

- The role of infrastructure
  - factorising needs supported with services
  - keeping the abstraction alive
- Mediating & governing infrastructures
  - services enabling / ruling agent interaction within the MAS environment / society
- Supporting trust creation & management
  - Enabling observation & traceability of agent action
  - Enforcing laws ruling agent / social environment interaction
  - Making laws inspectable & adaptable
- First class abstractions for mediating infrastructures
  - Coordination artifacts
  - Agent Coordination Contexts

Coordination Artifacts

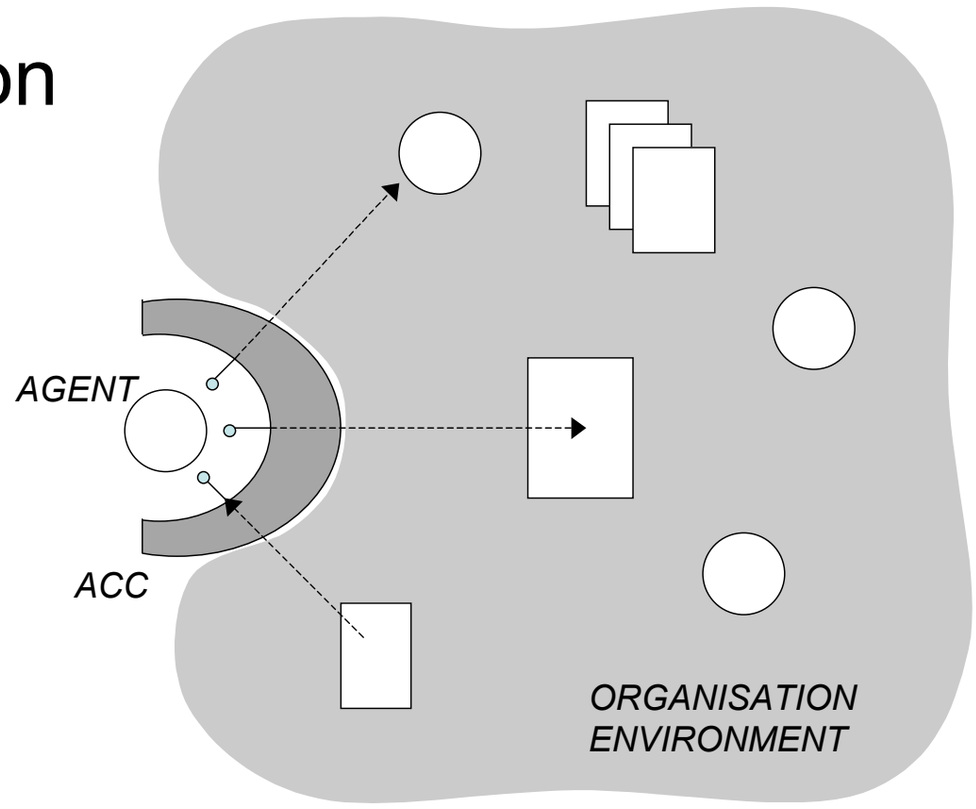
- Runtime abstractions specialised to provide a coordination service
  - constructive & normative aim
    - examples: blackboards, mailboxes, maps, schedulers, workflow engines, etc
  - abstract model
    - usage interface with operations
    - operating instructions
  - properties
    - encapsulation & specialisation
    - inspectability & controllability
    - malleability
- Toward infrastructures: TuCSoN
  - Tuple centres as coordination artifacts

Coordination Artifacts for Trust

- Embodiment of rules of the encounter
  - system-level trust
- Observation & traceability of agent interaction in social activity
  - individual-level + social Trust + Trust in societies
  - tracking society history
    - online and offline analysis for Trust evaluation
- Promoting constructive Trust
  - designer / user point of view
  - artifact inspectability / controllability / predictability for enhancing Trust in the system

Agent Coordination Contexts (ACCs)

- Representing the context of an agent inside an environment
- First class organisational abstraction
  - agent role(s)
- Interface to the org. environment
  - enabling actions and perceptions
- Role-based access control
  - ruling actions & perceptions
- Dynamic stages
  - ACC negotiation
  - ACC use
  - ACC update



Agent Coordination Contexts for Trust

- Linking Trust to organisation and security
  - ACC as contracts certifying individual (interaction) rights & duties
    - promoting Trust between individuals & organisations
  - ACC as interfaces applying role policies to avoid unwanted agent actions
  - system-level Trust

Conclusion

- The scope and notion of trust was re-framed / extended in the engineering perspective
  - users / engineers / individual agents / agent societies
- Infrastructure support for Trust creation and management
  - mediating / governing infrastructures
- Trust embedded in first class abstractions for coordination and organisation
  - coordination artifacts
  - Agent Coordination Contexts