

The complex facets of trust and reputation

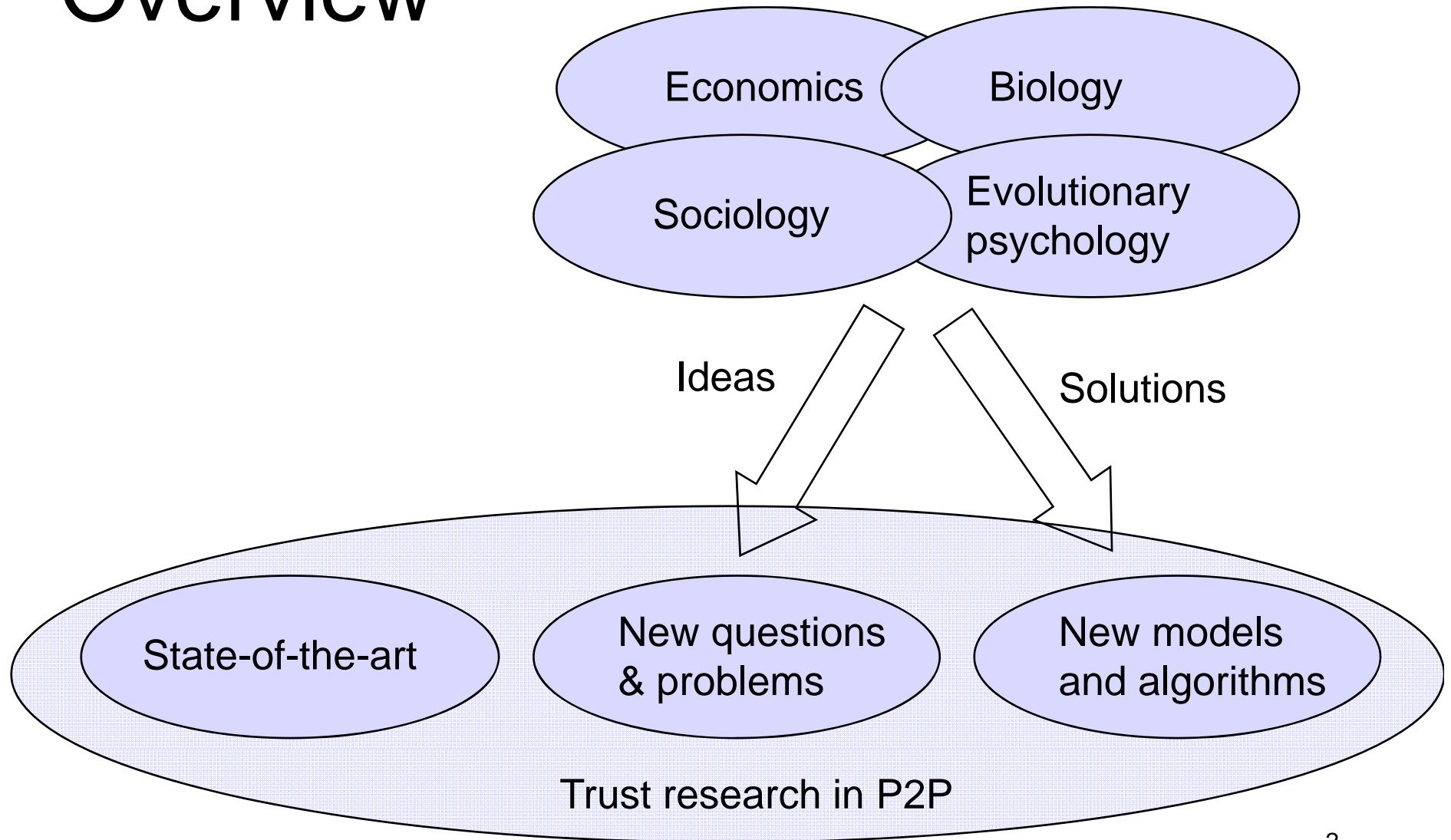


Karl Aberer (EPFL)
Wojciech Galuba (EPFL)



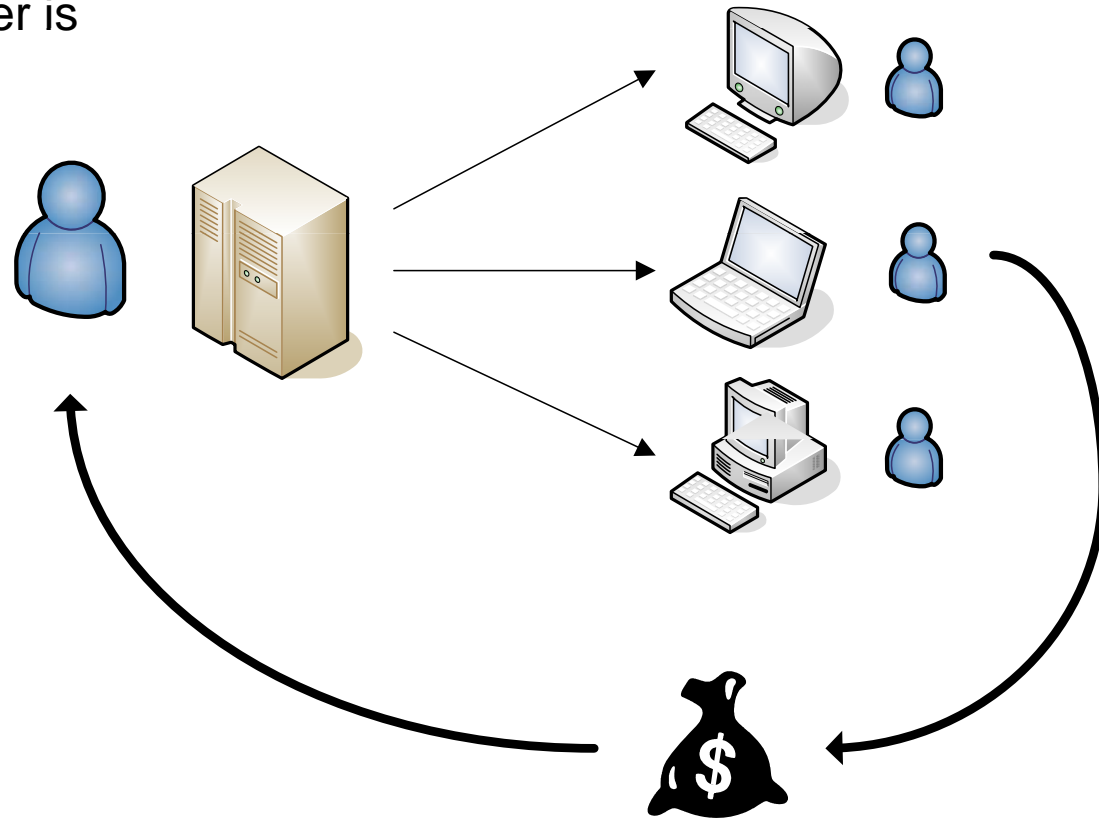
Zoran Despotovic (DoCoMo Euro-Labs)
Wolfgang Kellerer (DoCoMo Euro-Labs)

Overview



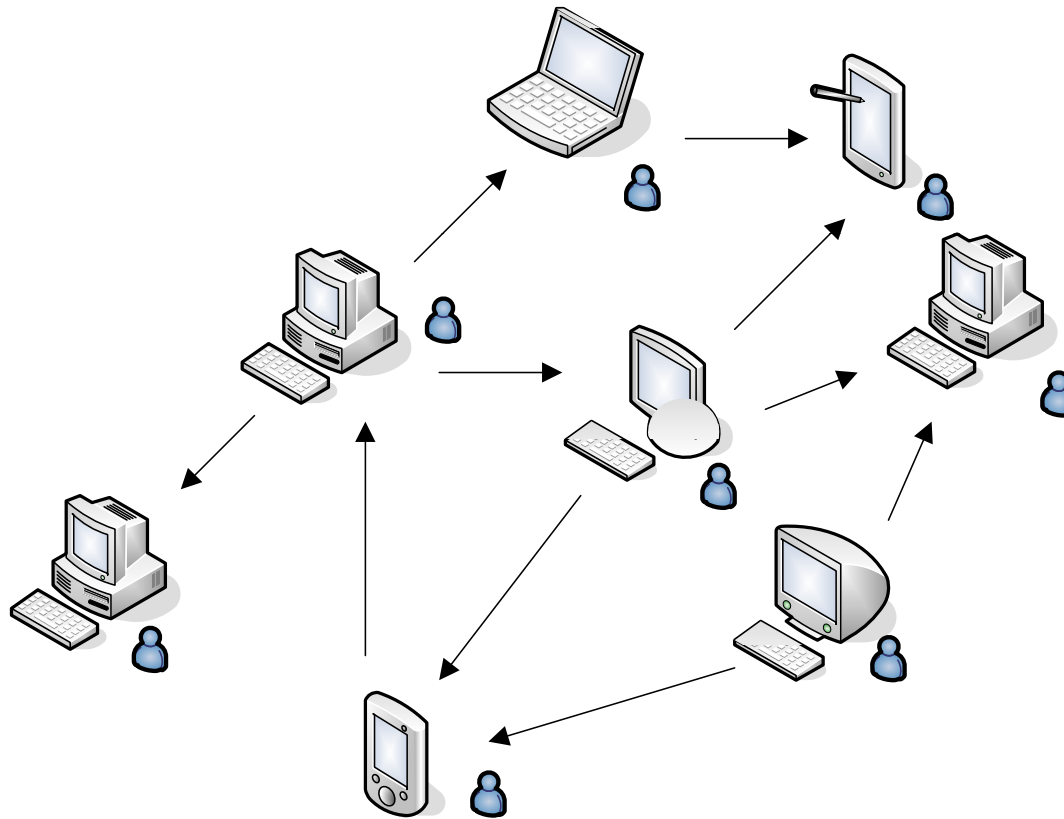
The two architectures

Client-server is sustainable



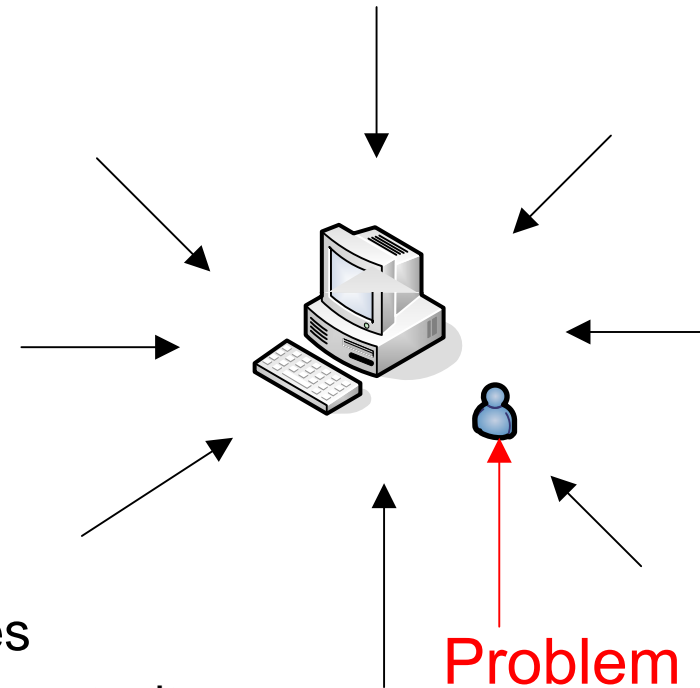
The two architectures

Peer-to-peer
Sustainable?



Peers (or rather users) are selfish

- Peer behavior might change
- To conserve:
 - Own bandwidth
 - freeriding in Gnutella
 - Own CPU cycles
- To exploit:
 - Greedily consume other's resources
 - ... before the others can use that opportunity
- **Cooperation needs to be built**
 - **How?**





Setting up the stage

- Pairs of peers **interact**
- Interacting peers face the **Prisoner's Dilemma**
 - Peers can either cooperate or defect
 - Example: transactions on eBay
- It's good to be able to predict if the other party will cooperate
- **Trust** – the extent to which a peer believes the other peer will cooperate
- **Trust model** – the way in which the above belief can be inferred
- Inference could be: if A cooperates with B then B will also cooperate with C with high probability
- Universally applied, the above inference constitutes **reputation**
 - The collective actions of the peer determine its reputation
- Hence: **reputation-based models of trust**



State-of-the-art approaches

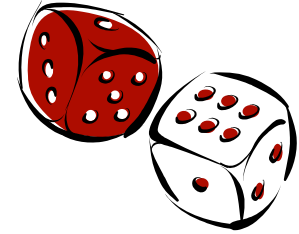
- Social networks
- Probabilistic (local) estimation
- Game theoretic approach
- Evolutionary approach



Social networks

- Interactions occur along the social links
- Social links carry history of interactions
- If you meet someone new you want to know if you can trust him/her
 - you ask your neighbors in the social network
- We make it recursive:
 - Node B wants to compute its trust in node A
 - We enumerate all the paths from A to B
 - Compute the trust values along the paths
 - Aggregate the computed values from all paths
- Social network approaches differ in how they enumerate, compute, aggregate etc.

Probabilistic estimation



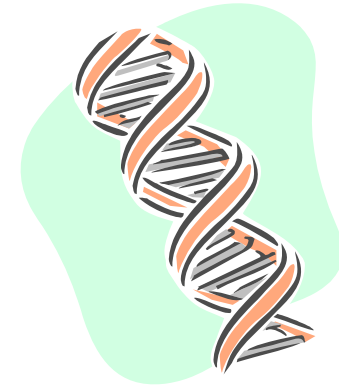
- The computed trust value in social networks hard to interpret
- We are more interested in the actual **probability of cooperation**
- Peers report cooperations and defections of their interaction partners
- Other peer can compute the probability of cooperation based on those reports
- Misreporting has to be considered

Game theoretic approach



- Perfect rationality assumed
- Not necessarily perfect information
 - Peers perceive the actions of other peers with error
 - Peers can make errors when executing their actions
- Utility functions of peers known
 - Nash equilibrium can be computed
 - Nash equilibrium determines the behavior of the peers

Evolutionary approach



- Population of PD-players
- Scores:
 - Each player has its score
 - Scores are updated based on the outcome of PD interaction
- Fitness = accumulated score
- Highest fitness players reproduce more frequently
- Evolutionary stability of a strategy:
 - population playing the strategy does not extinct under invasions of defectors
 - stronger: defectors are driven out of existence
- The winning strategy: **tit-for-tat**
 - Initially cooperate, punish defectors, forgive if cooperated



Intermission

- The current approaches make a number of simplifying assumptions
 - Or are highly stylized
- What if those assumptions are violated?
- Is it still possible to build a robust reputation system?
- How do the dynamics of reputation change?
- What are the winning behaviors?



Direct vs. indirect reciprocity

- Direct reciprocity: peers interact repeatedly
 - direct revenge and punishment possible
- Indirect reciprocity: one-shot interactions only (e.g. eBay)
 - If I scratch your back you should scratch someone else's
 - Peers need to exchange observations
- Tit-for-tat is the winner for direct reciprocity (Axelrod)
 - What is the winning strategy for indirect reciprocity?

The leading eight

- Ohtsuki et al.
- Information exchange via one bit label associated with each peer:
 - One bit publicly readable
 - Writable by all except the labeled
 - One bit is sufficient (Kandori et al.)
- There are 4096 strategies, only 8 are cooperative and evolutionary stable
- Analogy to tit-for-tat

action function:

GG	GB	BG	BB
C	D	C	*

assessment function:

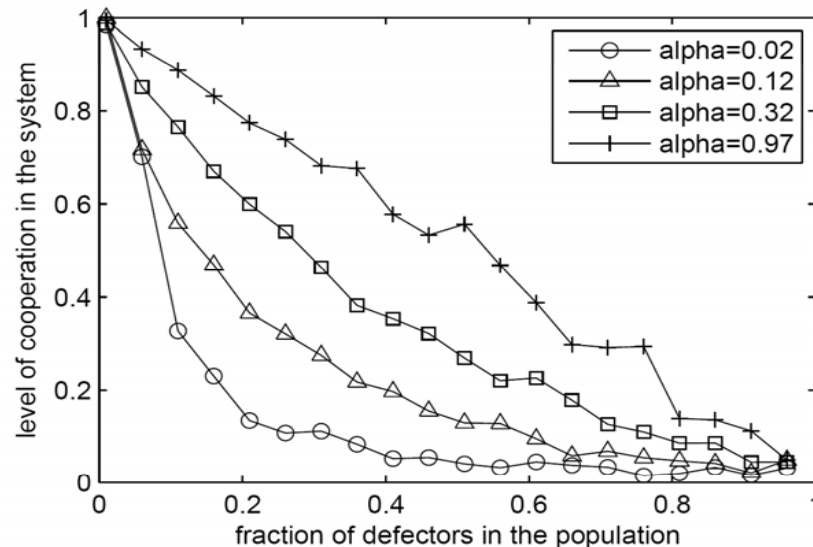
	GG	BG	GB	BB
C	G	*	G	*
D	B	G	B	*



Propagation of reputation information

- It is commonly assumed that propagation of reputation information is instantaneous
 - e.g. globally accessible one bit labels
- In reality:
 - Reputation propagation is delayed
 - Reputation update may not reach all nodes
- At least one piece of evidence from economics:
 - Delays lead to more efficient Nash equilibrium

Partial reputation propagation



- alpha – fraction of nodes to which reputation updates propagate
- It may not be necessary that all nodes receive the update

Bounded rationality



- Traditionally in game theory: perfect rationality
 - but Nash equilibria are NP-hard to compute
- The economists start leaning towards bounded rationality:
 - psychological evidence: we are not perfectly rational
 - we can be biased, use heuristics and we learn
 - bounded rationality explains empirical anomalies in economics (where unbounded rationality fails)
 - good decisions are costly, reliable information hard to get, computational power is scarce
- How can we incorporate bounded rationality into reputation models?



Behavioral evolution

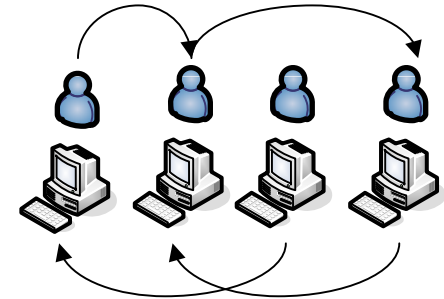
- Deciding on behavior costly? → Let's copy it from someone else.
- Two main ways of behavioral imitation in nature:
 - **payoff-based transmission** – imitate the highest gain behaviors
 - **conformist transmission** – if the majority has an iPod I'll also get one
- These two types of imitation can stabilize cooperation
 - according to evolutionary biologists (Boyd et.al.)



Evidence of behavioral evolution in P2P

- eMule – evolutionary arena:
 - Open source, modifications easy
 - <http://www.emule-mods.de> – various eMule versions with different behaviors
 - Selfish users modifying code
- Evolution example:
 - Nick thief – steals identity and reputation that goes with it
 - Anti-nick thief – adds cryptography to protect session identity
 - Anti-anti-nick thief – detects anti-nick thieves and bans them
- And much more:
 - Tit-for-tat naturally arose
 - Collusion by friend uploads (considered bad)
 - Stealth mode: only upload/download, no file announce

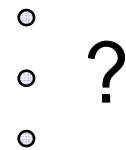
Evolutionary P2P?



- Richness of behaviors naturally arose
- Behavior imitated: mods propagate via user websites, discussion forums etc.
- Collective behavior fitness evaluation
- „Moral laws” spontaneously emerged
 - Mod blacklists appeared
- User goals are changing all the time
- How can we define the evolutionary stability in this case?
- **Social network of users cannot be considered separate from the P2P network**
 - How can we model this?
 - How much cooperation enforcement should we engineer into the system?
 - and how much of this should be left to the user community?

Second-order defection problem

- Cooperation enforced by reputation system
 - Reputation system has a distributed implementation
 - Nodes need to cooperate to sustain the reputation system
 - In particular: they need to exchange reputation info
 - This incurs additional cost
 - Second-order cooperation problem
- Usual practice when evaluating a reputation system:
 - Introduce second-order defectors (e.g. reputation liars)
 - See if first-order cooperation survives
 - Why never look at second-order cooperation sustainability?
- A solution: incentives
 - But this merely delegates the problem to another layer



Meta-reputation
Reputation
Base layer



How is SODP solved in nature?

- Biology and psychology:
 - Group selection – groups that use reputation are more cooperative hence more fit
 - Conformist transmission – if every one gossips I will too, it must be fun
 - Costly signaling – reputation propagation = the peacock's tail = advertising
- How can we re-use these solutions to deal with SODP in P2P?



Inhomogenous interactions

- Commonly assumed that interactions are equally probable for all pairs of nodes
 - But interaction topology has huge influence on behaviors in the equilibrium
 - Research on spatial EPD and on scale-free networks
- In P2P systems interactions are determined by:
 - overlay maintenance algorithms
 - file search algorithms etc.
- How can we characterize those interaction inhomogenieties?
- How do they limit the reputation system that we want to build?



Inhomogeneous interactions

- Interaction topology is dynamic:
 - what's more nodes decide who they want to interact with
 - part of the evolving node's behavior
 - some researchers have looked at network formation, i.e. what topologies arise in equilibrium
 - Wolinsky – economic models of network formation
 - Roughgarden – computer network formation
 - however strong game theoretic assumptions were made, e.g. utility functions are uniform and known
- Goyal – first models of simultaneously learning the actions and adapting the set of neighbors
- How can we build a reliable reputation system atop a constantly changing interaction topology?



Identity stability

- One of the most common criticisms of reputation systems:
 - How can they work without stable identities?
 - Identities are cheap easy to acquire or change
 - Vulnerability to Sybil and whitewashing attacks
- Identity can be made more costly by cryptography
- **Authentication is not a process returning a simple yes or no answer**



Identity stability

- More complex properties of identity:
 - Scope, e.g. In SSL links identities are provably stable for the duration of the connection
 - Partial perception, e.g. I may know that some node is from Dortmund, but not know which machine exactly
 - Lifetime
 - How to model this in a P2P system?
- **Identity verification domain \geq reputation domain**
- **What are the minimal requirements on identity verifiability needed to build a reputation system?**



Conclusions

- Up to now different approaches were making simplifying assumptions
 - To make the models analytically tractable
 - To narrow the parameter space for the simulation
- More realistic assumptions are needed
- Before we design a reputation system we need to consider:
 - Communication model
 - Computational constraints
 - Peer software dynamics
 - Interaction model
 - Identity model
 - Peer goal dynamics



Take home messages

- Distributed systems engineering
 - needs models that give predictions and guarantees
- In large-scale systems Byzantine fault tolerance is too expensive
 - Need to narrow down to a subset of faults: malicious and strategic node behavior




Conclusions

- What assumptions need to be made in each of these areas?
- How do these assumptions influence the performance of the potential reputation system?
- What is the minimum set of assumptions to make the reputation system possible at all?
 - Can we get some impossibility results?
- How can we re-use the existing solutions to common reputation system problems from biology, sociology and economics?



Applied soft-computing

- Combining knowledge and reasoning with uncertainty about:
 - Identity
 - Reputation information
 - QoS
 - Additional difficulty: this changes over time
- Evolutionary methods:
 - Models of peer behavior
 - Proving stability: „strategyproofness”



Before Q&A, a game ☺

- No communication is allowed before or during the game
- Each player writes his/her name and an integer from 0 to 100 on a piece of paper
- The average of all the numbers is computed
- The winner is the player whose number is the closest to $0.7 \times \text{average}$



The interpretation of the results

- The Nash equilibrium is: everyone submits zero and everyone wins
- Rationality is bounded
 - Only a limited number of game recursions can be simulated in the mind in a finite time
- You also need to estimate the rationality of other players
 - If you select zero and they don't, you most likely loose
- The average can be interpreted as a measurement of the rationality of the group
 - Or rather the group's estimate about its rationality



References

- „*The complex facets of reputation and trust*” Aberer, K. and Despotovic, Z. and Galuba, W. and Kellerer, W.

In „*Computational Intelligence, Theory and Application*” book chapter 29, Springer, to appear, 2006.