The Interactive Cooperation Tournament

How to Identify Opportunities for Selfish Behavior of Computational Entities[1]

Philipp Obreiter[1] and Birgitta König-Ries[2]

[1] Institute for Program Structures and Data Organization
Universität Karlsruhe (TH), 76128 Karlsruhe, Germany
obreiter@ipd.uni-karlsruhe.de
[2] Institute of Computer Science
Friedrich-Schiller-Universität Jena, 07743 Jena, Germany
koenig@informatik.uni-jena.de

Abstract. Distributed reputation systems are a self-organizing means of supporting trusting decisions. In general, the robustness of distributed reputation systems to misbehavior is evaluated by the means of computer based simulation. However, the fundamental issue arises of how to anticipate kinds of successful misbehavior. Existing work in this field approaches this issue in an ad-hoc manner. Therefore, in this paper, we propose a methodology that is based on interactive simulation with human subjects. The requirements for such interaction are discussed. We show how they are met by the Interactive Cooperation Tournament, a simulation environment for identifying promising counter-strategies to the distributed reputation system EviDirs which is showcased in our demo.

1 Introduction

In this paper, we propose a demo of our Interactive Cooperation Tournament (ICT). ICT is a simulation environment that provides realistic evaluations of the robustness of our distributed reputation system. We describe *why* systems like ICT are needed and *what* functionality they need to provide in order to facilitate the engineering of robust distributed reputation systems.

Context. If you look at computer systems, there is a clear trend away from closed monolithic systems towards self-organizing artificial societies composed of autonomous entities with no central control and no commonly trusted unit. Examples are peer-to-peer systems, open multi-agent systems, and ad hoc networks. All these systems have a number of characteristics in common: In order to achieve their individual goal, it is necessary for the entities in the system to cooperate. However, due to their autonomy, on the one hand, entities will only cooperate, if it is beneficial to them, but on the other hand, entities are able to cheat in the course of a cooperation. In order to avoid being cheated on, an entity will only cooperate with entities it trusts. Yet, trusting decisions

[1] The work done for this paper is funded by the German Research Community (DFG) in the context of the priority program (SPP) no. 1140.

can only be taken by an entity if it has formed its beliefs regarding the other entities based on prior experiences. The means of doing so are prescribed by the algorithms of the *distributed reputation system*. Several of such systems have been proposed in the past (e.g. [1–3]). The distributed reputation system of [3] (*EviDirs*) exhibits several desirable properties and, thus, builds the foundation of the remainder of this work. We distinguish between two types of computational entities. *Normative entities* adhere to the prescriptions of the distributed reputation system. In an application scenario, these entities are found on the devices of those human principals who make use of the original system software, as it has been distributed by system's initiator. On the other hand, *strategic entities* are not compelled to comply to the prescriptions and, thus, exhibit selfish behavior. This situation arises whenever human principals are able to tamper the original system software.

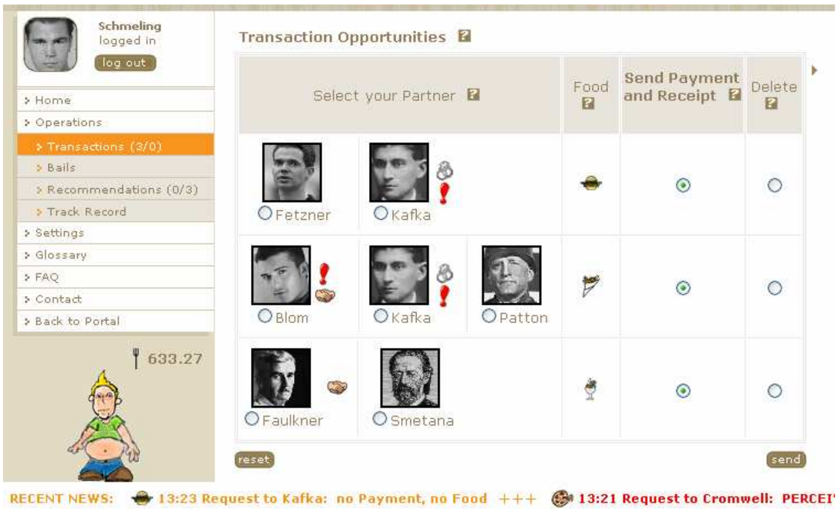
Problem. Analytic proofs are a viable means of displaying core properties of a distributed reputation system. Yet, their application is restricted to specific behavior within the overall system. This becomes apparent in their idealizing assumptions that have to be made in order to apply the methodology of game theory. Thus, they fail to capture and consider every opportunity of misbehavior by strategic entities. Consequently, a means of testing the robustness of the distributed reputation system is required. In analogy to the methods applied in evolutionary game theory, computer based *simulation* appears as a natural solution to both problems [4]. Even though this approach is viable, it poses a fundamental question to the designer and evaluator of distributed reputation systems: *Is it possible to anticipate how the system software is tampered and, if yes, what kind of tampering has to be anticipated?*

The state of the art regarding this question is as follows: The evaluator defines the counter-strategies to the system design according to his intuition. This approach suffers from two considerable drawbacks. First, the evaluator may overlook more intricate means of misbehavior. The existence of such means is probable due to the complexity of common distributed reputation systems. Second, the evaluator is, in general, also the designer of the distributed reputation system. Consequently, he might be induced to consider only those counter-strategies that his design is able to cope with. As a result of these drawbacks, we require a means of reliably identifying counter-strategies that should be included in the system's simulative evaluation.

Outline. In order to solve this problem, we propose the following approach: The simulation environment is made *interactive* such that human subjects may assume the role of certain entities. The simulation environment is built such that the human subjects are both able and motivated to find promising counter-strategies. In Section 2 the ensuing requirements and its implementation example for EviDirs is discussed. The obtained simulation environment ICT. It will be showcased in our demo.

2 The Interactive Cooperation Tournament (ICT)

In this section, we discuss the requirements that arise from our approach of interactively identifying promising counter-strategies. For each requirement, we point out how it can be implemented. In this regard, the ICT acts as a point of reference.



Select your Partner			Food	Send Payment and Receipt	Delete
○ Fetzner	○ Kafka			◉	○
○ Blom	○ Kafka	○ Patton		◉	○
○ Faulkner	○ Smetana			◉	○

RECENT NEWS: 13:23 Request to Kafka: no Payment, no Food +++ 13:21 Request to Cromwell: PERCEI'

Fig. 1. The user interface of the Interactive Cooperation Tournament

As a basic requirement, the human subjects have to be *informed* about the incidents that happen in the overall system. Furthermore, mere information is not enough since it has to be presented in a user-friendly manner. Only if this is achieved, the human subjects are able to intuitively grasp the context of their behavioral decisions and choose their respective strategies accordingly. Figure 1 illustrates how the ICT implements this requirement. Each participating entity is assigned an avatar so that the human subjects are more likely to recognize them and remember their prior behavior. The avatars and the labeling of the entities are assigned randomly for each tournament so that the human subjects do not know which entity is controlled by the simulation environment and which are not. This corresponds to the situation in real system in which the entities do not know who is normative and who is strategic. Each human subject is able to access the complete record of the experiences his entity has made. That information is aggregated and displayed by additional icons nearby the avatars.

A further requirement consists of the *accessibility* of the simulation environment. The human subjects do not have to be experts of the distributed reputation system EviDirs in order to participate. Only by this means, the number of potential human subjects is not restricted and, hence, a wide range of counter-strategies can be obtained. The ICT takes this requirement into account by providing a tutorial, a glossary and a forum in which system-specific questions are debated. As a further assistance for the human subjects, the simulation environment may control behavioral aspects (e.g., the more intricate recommendation behavior) that novice subjects are not able to cope with.

The third requirement refers to the *motivation* of the human subjects. In real application scenarios, tampered system software is used in order to reduce one's own costs or enhance one's own benefits of participating to the system. As a consequence, we have to anticipate counter-strategies that aim at maximizing the individual utility of the entities that follow them. The ICT makes use calories as the metaphor of individual utility. On the one hand, the cost category of transactions is illustrated by adequate food icons (a pizza has more calories than an apple...). On the other hand, one's own individual utility is visualized by a guy on the lower left corner: the fatter he is the more successful the human subject has performed. A further source of motivation is the policy to contact the most successful subject after termination of the tournament and to ask for the counter-strategy he has followed. By this means, the evaluator obtains information about the kind of counter-strategies that are most successful and, thus, have to be anticipated. According to our experiences, the human subjects are able to express the basic principles of the strategy (or strategies) they followed. By additionally consulting the simulation log, the evaluator is able to define and parameterize the identified counter-strategies.

For the demo, we run an instance of the tournament and allow visitors to explore themselves how it works and find out how good the strategies are that they follow.

3 Conclusion

Distributed reputation systems are a self-organizing means of supporting trusting decisions. In order to simulate such systems, we have to be able to anticipate successful (and thus likely) counter-strategies that could be pursued by misbehaving entities. In this paper, we have proposed a methodology that is based on interactive simulation with human subjects. The requirements for such interaction are threefold: The simulation environment has to inform the human subjects appropriately about the system's incidents, it has to be accessible to a wide range of potential subjects and, finally, the human subjects have to be motivated to maximize the individual utility of the respective entity they control. Furthermore, we have shown how these requirements are met by the ICT, an interactive simulation environment of the distributed reputation system EviDirs. As a result of meeting the requirements, the participating human subjects have identified promising counter-strategies to EviDirs in a hands-on manner.

In the future, similar interactive simulation environments have to be developed for other distributed reputation systems. This necessity arises from the fundamental problem of anticipating realistic means of misbehavior. Even though the development of adequate simulation environments or, if possible, of a generic simulation environment is an intricate task, it provides the only means of solving this problem and, thus, credibly testing the robustness of arbitrary distributed reputation systems. The discussion of this paper provides the basis for such future work.

References

1. Despotovic, Z., Aberer, K.: A probabilistic approach to predict peers' performance in P2P networks. In: 8th Intl Workshop on Cooperative Information Agents (CIA'04). (2004)
2. Liu, J., Issarny, V.: Enhanced reputation mechanism for mobile ad hoc networks. In: Second International Conference on Trust Management (iTrust'04), Oxford, UK, Springer LNCS 2995 (2004) 48–62
3. Obreiter, P., König-Ries, B.: A new view on normativeness in distributed reputation systems – beyond behavioral beliefs. In: Fourth Workshop on Agents and Peer-to-Peer Computing (AP2PC'05), Utrecht, Niederlande (2005)
4. Axelrod, R.: The Evolution of Cooperation. Basic Books (1984)