Chapter 1

Introduction

If you took the most ardent revolutionary,
vested him in absolute power, within a year
he would be worse than the Czar himself.

Mikhail Bakunin

The field of Distributed Artificial Intelligence (DAI) studies how different processes can work together in order to solve specific goals. Multiagent Sytems (MAS) are a subset of DAI systems in which the different processes (a.k.a. agents) have been developed by different entities. Therefore, each agent may have its own goals which it wants to satisfy by interacting with other agents in the MAS.

An agent chooses to interact with other agents because it is not able to achieve its goals on its own. Nonetheless, interaction with other agents does not guarantee that it will achieve its goals. The agent must find those agents that can help it achieve its goals and interact with them. In an interaction each agent has some expectations as to how the other agents will behave. These expectations can be exchanged and agreed to before the interaction takes place (*e.g.,* as in signing a contract), they can be fixed as part of the environment in which the interaction takes place (*i.e.,* by having the system designer specify the norms that will govern interactions), or none of the above, thus having agents interact blindly with one another allowing expected behaviours to be learned through time. Having agents in an interaction know one anothers expectations does not guarantee that they will be satisfied with the interaction results. Agents need mechanisms to enforce the behaviours they expect from others.

Humans have had to deal with these same issues, thus it is an interesting exercise to see what solutions have been proposed for humans. According to Taylor [Taylor, 1982], enforcement of expected behaviours has been achieved in primitive human societies through techniques that can be categorised through one or more of the following:

- Persuasion - where an agent modifies the beliefs of other agents through reasoning, so that they will believe that following the expected behaviour is preferable (*e.g.,* John persuades Peter to drive on the right hand side of the road by explaining why doing so avoids collisions).

- Authority - where an agent can modify the beliefs of other agents through its endorsement of the expected behaviour without giving reasons about why the behaviour is best (*e.g.,* Peter drives on the right hand side of the road because John, which is a government official, told him it is best).

- Power - where an agent can execute actions that change the probabilities of other agents achieving their goals (*e.g.,* Peter drives on the right hand side of the road to avoid being fined by John, which is a police officer).

- Physical Constraint - where an agent can bring about actions that impede other agents from continuing to interact (*e.g.,* Peter drives on the right hand side of the highway because it is impossible for him to drive on the left hand side, since John, the engineer in charge of building the highway, placed a barrier dividing the two sides).

Techniques that classify under persuasion involve a high degree of cognitive capabilities. Those classifying as authoritative either expect agents to have capabilities through which they can model each others degree of authority, or they must be hardcoded into their instincts. On the other hand, techniques based on power and physical constraint involve dependencies amongst agents which can be used as rewards or sanctions.

When human societies are looked upon for examples of enforcement techniques based on physical constraint, one starts by identifying those physical characteristics common to humans that can be used to sanction and reward them. These common characteristics are the fact that humans feel emotions, such as pain, pleasure, shame, and loneliness. All these emotions can be used in order to get a human to act in a certain way. For example, the threat of inflicting pain has been a common enforcement technique in many human societies. Nonetheless, as of now, most of these emotions are not present in artificial agents. One could make an exception by stretching the meaning of loneliness by matching it to a sense of gregariousness. In a way, artificial agents want to be in "company" of other agents, since they need them in order to achieve their goals. Nonetheless, the enforcement technique that uses the need to interact with other agents can also be classified under the power category.

In order to identify power-based enforcement techniques, one has to identify the resources that the agents need in order to achieve their goals. Using human societies again as an example, we observe that their basic needs to achieve goals are: energy, time, physical resources, and information. These can all be transferred in some way or other from one human to another. A human's energy and time can be used to help another human achieve its goal, *e.g.,* Tom can use his strength and time to build Anne's new cupboard. Physical resources are limited and access to them can be granted to other humans, *e.g.,* Britney can lend her

car to Alan so that he can get to work. Finally, information can be spread to others, *e.g.,* Cecile can advise Diego on which are the best weather conditions and routes to reach the Aconcagua summit. The threat of resource access denial is a power-based enforcement technique.

In order to find techniques that work for artificial software agents, one must start by identifying the resources they need to achieve goals. Surprisingly, these are the same as those seen for humans: energy, time, physical resources, and information. An agent needs energy as a power source in order to run in a computer. It also needs time for its computations, and information as an input for them. Finally, it also needs physical resources: CPU, memory, and bandwidth. Out of these physical resources it is only bandwidth that can be transferred. Unless we are dealing with mobile agents, which we are not in this thesis.

The work in this thesis describes power-based enforcement techniques where bandwidth is the physical resource used as the incentive to achieve expected behaviour. By blocking access to the resources needed in order to interact, an agent can enforce behaviours on others. This thesis shows how blockage from bandwidth usage through a network of agents impacts the ability to interact of agents that do not exhibit the expected behaviours. Models for MAS structured as networks are described where these blocking techniques can be applied, and the impact of such enforcement techniques are shown analytically and experimentally. Structuring a MAS as a network is a natural phenomenon since the advent of the internet, which is a network of networks through which humans communicate. It is becoming a widespread occurrence ever since the appearance of social networks. Although these are virtual networks in a centralised application scenario.

1.1 Motivation

Many distributed systems have appeared since the internet became a well-known technology. Through these systems many users around the globe come together to interact with each other and achieve certain goals. At first these systems where closed in many ways. Access was limited, and so were the actions available to users. Under these conditions is was fairly easy to get users to exhibit the expected behaviour. This was achieved in a centralised manner by the system designers which enforced the behaviours they wanted.

As the users grew in numbers and the technology evolved to allow more personalisation, it became harder for the system designers to enforce the behaviours that would suit all users. In order to satisfy users better, distributed system designers would have to give the users enforcement capabilities of their own and allow expected behaviour to emerge with time other than engineer it beforehand.

In today's systems, the only way users can enforce their expected behaviours is by deciding not to interact with those that they believe will not satisfy their expectations. Either because they have not done so in the past, or because they have come to know about previous interactions and realised that they have incompatible expectations.

Another technique for enforcement is to get others not to interact with a specific agent. This is done indirectly in current systems by publishing interaction feedback so that others with similar expectations choose not to interact with the given agent. Current online systems have incorporated technologies that aid in this gossip gathering process in order to assess the probability that an agent will satisfy specific expectations. These technologies are known as Trust Managements Frameworks (TMF). Nonetheless, TMFs do not empower agents with new enforcement methods per se. They just give agents better information tools so that they can decide when to interact with other agents.

The motivation in this thesis is to find new methods through which the probability of having a satisfactory interaction is increased. This is achieved by enforcing the expected behaviours. Particularly, we are interested in those enforcement techniques that are totally distributed, since centralised techniques would not be as robust and may suffer bottleneck problems. By distributed techniques we mean those that can be aplied by all agents in the system while still allowing agents to have personal expectations and to have their own policies as to how strictly they want to enforce. Furthermore, these policies must take into account that other agents will try to avoid the sanctions being applied. Therefore, it is very important that the techniques provided are robust to these evasion techniques.

A system in which these distributed enforcement techniques are available should allow different communities to emerge in an efficient manner. Each of these communities would be formed by a set of agents whose expectations on the behaviours of others are compatible. Something like this is already happening in the Internet, where communities with different interests have formed. The difference being that, when enforcement techniques are provided through technology, they are either not efficient or only available to a selected few.

1.2 Contribution

The main idea underlying all our contributions is to structure a multiagent system through a social network. In small human communities, such as indigenous tribes or villages, all members know each other personally and know what to expect from each other from the outcome of many previous interactions. As communities grow in size, it is hard for those that conform them to know each other as is the case of cities or groups of villages scattered through an area. In such cases the lack of information about others can be overcome through third parties. By depending on these third parties for the information, they are being given power which can be used as an enforcement technique.

Let us illustrate this through an example. Albert, which has lived in the small town of Aberdale all his life working at his uncle Bill's farm, wants to move to the neighbouring town of Springfield where he plans to work at another farm. Albert knows no one at Springfield, and none of the farmers at Springfield have any knowledge about Albert's capabilities at farm work. Notwithstanding, Albert has a recommendation letter from his uncle Bill, which is well known by

Charlie (a Springfield farmer) from previous deals at different farmer markets. Since Bill vouches for Albert, Charlie is willing to employ Albert at his farm. The interaction in this example took place because of Bill's vouching through the recommendation letter. Had Albert done something in the past that did not satisfy Bill, he would have been able to deny Albert a recommendation letter, thus lowering Albert's possibilities of working at Springfield.

Our contribution consists of two protocols for interaction bootstrap that force agents to depend on their contacts in the social network, thus giving agents a degree of power over their contacts which can be used through different enforcement techniques we have proposed. We also provide a mathematical model for multiagent systems structured as social networks, and we give some analytical results showing that a decrease in unsatisfactory interactions can be achieved under certain conditions. Nonetheless, the model provided allows agents many degrees of freedom in their behaviour. In the analytical exploration, we narrowed this freedom through strict assumptions on their behaviour and their expectations. In order to make up for this, we have also realised experiments in more complex scenarios.

Through the experiments, we have tested to what extent the proposed enforcement techniques can increase the ratio of satisfactory interactions. Furthermore we also tested how well specific approaches fared against different enforcement evasion techniques that are known to be used by malicious agents in current systems.

1.3 Thesis Structure

Chapter 2 surveys the current state of the art. In that chapter the relationship to other research fields is studied, and so are recent developments in the field of multiagent systems, specifically those that deal with trust and reputation management and those that deal with group formation. Furthermore, other distributed systems where the proposed enforcement techniques can be used are also reviewed.

Then Chapter 3 defines the experimental methodology that has been followed throughout the experiments. The methodology has guided the experiment design, the data gathering, and the subsequent statistical analysis of the experimental data results from which we have validated the original hypothesis. A chapter has been dedicated to explaining the methodology since it has been used for the experiments in Chapters 4 and 5.

Chapters 4 and 5 are the core part of the thesis. These chapters provide the interaction bootstrap protocols through which agents can enforce the expected behaviours on others. Chapter 4 describes a protocol for interaction bootstrap in which an agent searches for an interaction partner which is *not known* before the protocol starts. This chapter provides the first set of enforcement techniques which can be embedded into the interaction bootstrap protocols. The definition of a satisfactory interaction in this chapter is engineered through norms and shared by all. Therefore, all agents have the same definition of a satisfactory

interaction which is objectively verifiable by all. Furthermore, analytical results are provided which give an upper bound to the number of unsatisfactory (*i.e.,* norm violating) interactions when specific conditions are met in the multiagent network. Finally, the results of experiments are provided which show the outcomes in less restrictive scenarios.

In Chapter 5 a second interaction bootstrap protocol is provided. In this case the partner with which the agent wants to interact is *known* from the beginning of the protocol. This difference allows for a new enforcement technique which can be added to those provided in Chapter 4. Nonetheless, the definition of a satisfactory interaction in this chapter is made subjective, *i.e.,* each agent has its own definition and these definitions are not necessarily known by other agents. Therefore, the analytical results in Chapter 4 no longer hold and a new analysis has been realised. Finally, the updated set of enforcement techniques are tested against other enforcement mechanisms through experiments. These experiments test the reduction of unsatisfactory interactions, and also test whether the enforcement techniques are robust against adversarial behaviours by malicious agents.

Finally, Chapter 6 wraps up the thesis by providing the limitations of the approach, and how they may be tackled. It also provides some examples as to how the enforcement techniques can be applied to currently functioning systems.