Simulation in Virtual Knowledge Communities

Jacques CALMET
calmet@ira.uka.de
IAKS, University of Karlsruhe
Karlsruhe, D-76131, Germany

and

Julien SUBERCAZE
subercaze.julien@gmail.com
IAKS, University of Karlsruhe
Karlsruhe, D-76131, Germany

and

Pierre MARET
pierre.maret@insa-lyon.fr
LIRIS (CNRS UMR 5205), INSA de Lyon
Villeurbanne, F-69621, France

ABSTRACT

The information society has reached a new era of communication and knowledge management. Virtual communities are emerging as a requirement in this new information distribution scheme. We investigate the concept of virtual knowledge communities using an agent-based abstraction. In this field, there is a need for evaluating the performances and the quality of a global system resulting from the behavior of distributed and autonomous entities. We provide a theoretical background for simulation as well as an implementation of a simulator based on our prototype for virtual knowledge communities.

Keywords: Virtual knowledge communities, Multi-agent-system, Simulation, Effective Knowledge Distribution

1. INTRODUCTION

In many domains, knowledge management has became a very important part of our society. The virtual knowledge community (VKC) is an interesting feature of knowledge management, based on interactions among different entities. There are several definitions for VKC. We can cite two of them: 'Groups of people with similar interests who communicate and interact in an online environment'[2] and 'A virtual community is a group whose members are connected by means of information technologies'[1]. VKCs are defined upon three ground concepts: different entities (the members), similar interests of the members, and electronic communication channels. As defined above, VKCs cover all the common knowledge exchange on the internet such as email, forum, newsgroup and many more. Applications in business are also very broad. In the relatively new concept of virtual organisation ('A virtual organisation is a collection of individuals, companies or organisations who have agreed to work together to achieve a purpose') [10], the knowledge management is the keystone for the success of the organisation. The VO can be seen as virtual communities with a commercial collaborative goal.

Many systems have been developed to support virtual knowledge communities and heterogeneous sources [9]. The common natural approach for developing virtual knowledge communities is the Multi-Agent System paradigm (MAS) [11]. Indeed, MAS perfectly meets the requirements to model virtual knowledge communities, considering the former definitions. Following the AOA paradigm [4], a first prototype was developed [7], followed by a second version using a decentralized approach for the communication scheme. The primary goal of our work was to introduce the concept of scenarios in the system, so that we can represent different situations and analyze the behavior of the system.

This paper is structured as follows. In a first part we describe our approach on simulation for VKC, then possible formal definitions are given. In the next part we introduce the measure of the efficiency of the knowledge distribution and we give experimental results. Then, we shortly describe the implementation of our simulator. Finally, we conclude by providing directions for future research as well as final remarks.

2. SIMULATION

To evaluate quantitative properties of MASs, three formal models are widely used, the process algebras [3],

automata and Petri nets [8]. In our case we do not want to simulate the behavior of a non implemented system, but to simulate different scenarios of implemented VKC prototypes. For example, we have some broker agents, some reseller agents and the client agent. Brokers want to provide the service to the resellers for the best price to the clients. If we assume the existence of an up-to-date yellow page directory a broker will have no difficulty in finding the resellers. If this directory does not exist, how will a broker be able to discover new resellers and their services? How efficient will be a decentralized solution, and which solution is the best suited to a given system ? The feature we want to evaluate is the network structure of the MAS. Other features of the system that can be evaluated will be discussed further. First of all we define the process of simulation.
We decompose a simulation process into three parts:

1. Design of the scenario.
2. Execution of the scenario and collect of the simulation data.
3. Analysis of the data.

In our example, one scenario can be the following: a centralized MAS, n agents belonging to the following categories: broker, reseller and client. Our interest was to evaluate the communication cost of the system, so we want to collect the exchanged messages. We commonly assumed that the agents have a BDI structure, so the achievement of their goals is also an interesting data. The analysis should provide an efficiency measure of the system, that can then be compared with others. This short description of a simulation leads us to a formal definition of a simulation.

Design of the scenario

We define a scenario as a tuple (S, A, C, Z, R, T).

- S is the system in which the scenario takes place. It can feature a MAS framework.
- A is the set of the agents.
- C is the set of the agent's classes. In the former example, there were three classes, the clients, brokers and resellers. Using the AOA paradigm, agents that belong to the same class share the same decision mechanism code and have their own knowledge.
- Z is the set of actions. From an automata point of view, an action leads to the change of the agent's internal state. Each agent class has its own actions. $Z = \bigcup_{c \in C} ActionsOf(c)$
- R is the set of relationships. For example if we want to modelize the corporate knowledge management, we have to define hierarchical relationships. The degree of complexity of the relationships will clearly depends on the system that one is trying to modelize. In a more general social context, we can define all the relationships that the human being knows, like friendship, love, hate etc. At the current state of the prototype, this concept of relationships does not exist and we will not discuss it any more, except in the conclusion. This set can be left as empty.
- $T \in R_*^+$ is the duration of the simulation. It can be finite or not, and cannot be equal to zero.

Execution of the scenario and collection of the simulation data

The execution of the scenario depends on the platform and on the structure of the selected MAS. This point will not be investigated in this paper, because it is strictly specific to a given system. A more interesting point is how to collect the data of the simulation. First of all we introduce the common class of data that are collectable on such a MAS, then we will discuss the analyse . We assume agents having a BDI architecture, exchanging messages over an electronic communication channel. We can define three classes of data:

1. Communication data: the exchanged messages
2. Satisfaction of the agents
3. Knowledge of the agents

To collect the exchanged messages, one may use a central agent that receive and store a copy of each message or rely eventually on the MAS framework [5]. Messages can be stored through the form: $(sender, receiver, content, timestamp)$. The satisfaction of an agent can be defined as a function that indicates whether the agent has achieved its goal. We propose to use a function $f : State \rightarrow [0, 1]$ where state is the current state of the agent. In the next section we use the time as parameter rather than the state. As for messages, a central entity can store the evolution of the internal knowledge of the agents. For a large scale system, a distributed solution [6] should be prefered. We will not handle this issue in this paper.

Analysis of the data

A relevant analysis of the data should provide measurements of the system efficiency. We introduce here some functions that could be used to build complex efficiency evaluation functions. There are no general solutions since the analysis is user and system dependent. One may privilege an aspect of the system and have no interest in the others.

Each of the former defined data classes provides us useful information. We aim at building relevant indicators of the system using the available informations. We define the function $Messages(t)$ that returns the number of exchanged messages until time t has elapsed. The global system satisfaction gives us information on the goal achievement of the agents at time t:

$$achievement(t) = \frac{\sum_{a \in A} f_a(t)}{|A|} \quad (1)$$

Combining those two functions we can define a more interesting function which is the ratio between the global system achievement and the number of exchanged message: $\frac{achievement(t)}{Messages(t)}$.

It provides the measure of the evolution of the communication efficiency over the time. On one hand, this measure is relevant for centralized systems, where the overhead due to the agent discovering feature is very low. On the other hand, regarding peer to peer systems (with a long peer searching phase), we will prefere to give a weight on the messages depending on the time. In the next section we will introduce another relevant measure for MAS, the Effective Knowledge Distribution (EKD). We provide then experimental results.

3. EFFECTIVE KNOWLEDGE DISTRIBUTION

In a virtual knowledge community system, members (i.e. agents) share information. We want to measure how the knowledge is distributed among the members along the time. At time t the set of knowledge in the system will be denoted $K_{M,t}$. We denote $K_{i,t}$ the knowledge owned by the member $i \in A$, A being the set of members. A naive definition of the knowledge distribution $d(t)$ is then:

$$d(t) = \frac{\sum_{i \in A} \frac{|K_{i,t}|}{|K_{A,t}|}}{|A|} \quad (2)$$

This definition seems to meet the requirements for such a measure: $d(t)$ is equal to one when each member owns the whole set of knowledge, and it tends to zero when only few members own the knowledge. However this definition does not give a relevant measure of the knowledge distribution within the concept of scenario that we introduced in the former section. Within a social system, agents can dynamically enter and leave, so that the set of knowledge into the system dynamically changes. In order to better handle the problem of knowledge distribution, we have to consider the set of knowledge that was available during the scenario and not only at time t. Actually, reusing the definition of section 2, this set is equal to the union of the knowledge of each agent. As agents may join or leave during the simulation, we have to consider it over the duration : $K_{global} = \bigcup_{i \in A, 0<t<T} K_{i,t}$. We have also to consider that each member does not have interest in the whole knowledge, but only in a part of it (depending on its goal). We assume that each agent owns an interest-evaluation function that indicates its interest in a piece of knowledge. We define this function as $f_i : K_{global} \rightarrow [0, 1]$ where $f(k) = 0$ if the agent i has no interest in the knowledge and $f(k) = 1$ when the interest is maximum. As an example, in the next section we will use a function that takes only those two values.

We can now define the Effective Knowledge Distribution:

$$EKD(t) = \frac{\sum_{i \in A} \frac{|f(K_{i,t}|}{|K_{global}|}}{|A|} \quad (3)$$

The EKD ranges between 0 and 1. It's a relevant indicator for the knowledge propagation in a virtual communities system.

Experimental results

We use a simulation environment to measure the evolution of the Effective Knowledge Distribution. We simulate different agents' behaviour to observe the evolution of the distribution over the time. We assume that each agent has only one topic of interest, and that the knowledge is entirely present at the initialisation of the system (all agents are present at the beginning).

The following interest-evaluation function was used for each agent:

$$f_i(k) = \begin{cases} 0 & \text{if } i \text{ has interest in k} \\ 1 & \text{if } i \text{ has no interest in k} \end{cases} \quad (4)$$

The figure 1 shows the EKD for the three following scenarios. Similarly we computed the ratio between achievement and exchanged messages. The mean achievement of the system at time t is defined as follows:

$$achievement(t) = \frac{\sum_{i \in A} \frac{f_i(K_{i,t})}{interest_i(K_{global})}}{|A|} \quad (5)$$

$interest_i(K)$ is the number of knowledge instances meeting i's interest among the knowledge set K. The figure 2 displays the results.

Three basics scenarios were defined, using probabilistic behaviour of the agents and an iteration principle. At each iteration, each agent can exchange full request with one other agent (ask for a knowledge matching within their interest, receive the answer):

1. The agents send randomly knowledge to other agent, without regarding whether they have interest in it. (The dummy system.)

2. The agents ask randomly for knowledge matching their interest but without knowing if the agents they ask own such a knowledge.

3. The agents ask only agents owning knowledge matching their interest.

For the two first scenarios, *randomly* means that each agent has an egal probability to be chosen and queried. The results were computed using 1000 agents, sharing 2000 instances of knowledge in 20 topics (100 instances per topic) over 2000 iterations (Figures 1 and 2). A $t = 0$, agents may share knowledge that do not belong to their interest but they are only looking for knowledge matching their interest. The third system appears clearly as the most efficient in this simulation. It simulates the common centralized structure with an up-to-date yellow page directory: agents use the directory to find others agents that have knowledge about the same interest.

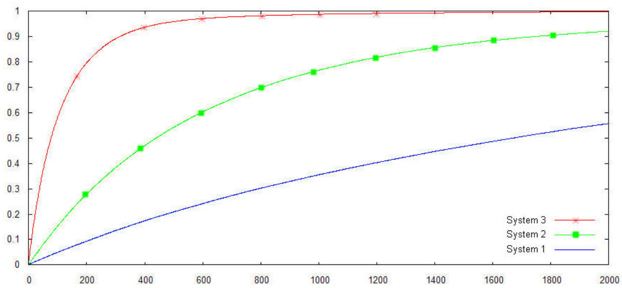

Figure 1. Effective knowledge distribution

4. IMPLEMENTATION

In the scope of our research, we developed several versions of a system of VKC [7]. The first version provided a centralized communication scheme using a yellow page directory (third example in section 2). The second version was a decentralized version of the VKC system. The ground idea was to be able to compare the behaviour of

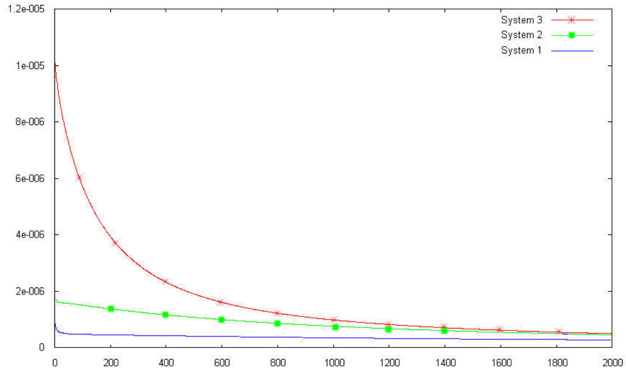

Figure 2. Achievement over exchanged messages

both systems in different situations. We aim at observing the behaviour of the system and at measuring the efficiency of the system in terms of communication cost, of efficient of the knowledge distribution. These measurements provide a good view on the efficiency of a system. Each situation requires to evaluate the efficiency. In a very large scaled system, a very efficient communication will be preferred, disregarding the knowledge distribution. On the other hand, for small or middle sized systems, a good knowledge distribution will be preferred. In order to represent the different situations that can be handled by systems, we introduced the concept of scenario. We implemented a scenario editor which is system indepedent and which enables the description of events, durations, etc. A GUI (Figure 3) was developed to provide a useful interface for this scenario editor.

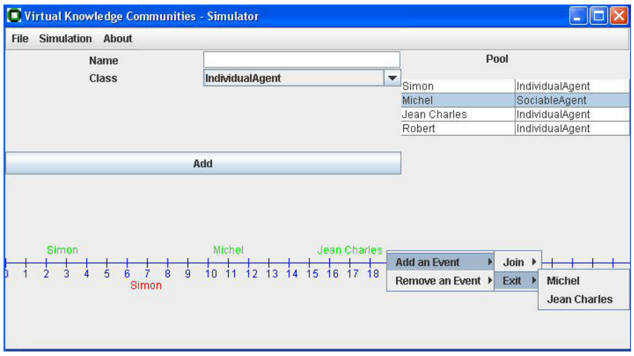

Figure 3. Scenario designer UI

5. CONCLUSION

A formal approach of simulation in the scope of VKC was given. We introduced formally the concept of scenario, we discussed the nature of the data and we introduced basic measurements for the global system. In this approach we mainly dealt with the analysis of the data / knowledge exchanges within MAS. One of the issues now is to address large-scaled systems. Since these systems can not be observed globally, there is a need for developing local methods to evaluate the efficiency of a MAS, particularly in the scope of VKC.

REFERENCES

1. Experts archive question. http://experts.about.com/e/v/vi/Virtual_community.htm.
2. John Beech and Simon Chadwick. http://www.booksites.net/download/chadwickbeech/, 2004.
3. J.A. Bergstra, A. Ponse, and S.A. Smolka, editors. *Handbook of Process Algebra*. North-Holland, 2001.
4. Jacques Calmet, Pierre Maret, and Regine Endsuleit. Agent-oriented abstraction. In *RACSAM (Revista Real Academia de Ciencias, Serie A de Matematicas Vol. 98 (1)*, December 2004.
5. Steven P. Fonseca, Martin L. Griss, and Reed Letsinger. Agent behavior architectures a MAS framework comparison. In *AAMAS*, pages 86–87, 2002.
6. Melanie Gnasa, Sascha Alda, Jasmin Grigull, and Armin B. Cremers. Towards virtual knowledge communities in peer-to-peer networks. In James P. Callan, Fabio Crestani, and Mark Sanderson, editors, *SIGIR Workshop: Distributed Multimedia Information Retrieval*, volume 2924 of *Lecture Notes in Computer Science*, pages 143–155. Springer, 2003.
7. Pierre Maret, Mark Hammond, and Jacques Calmet. Virtual knowledge communities for corporate knowledge issues. In *Proceedings of ESAW 04, LNCS 3451*, pages 33–44, 2004.
8. C. A. Petri. *Kommunikation mit Automaten*. PhD thesis, Institut für instrumentelle Mathematik, Bonn, 1962.
9. Alun D. Preece, Kit ying Hui, W. A. Gray, Philippe Marti, Trevor J. M. Bench-Capon, Dean M. Jones, and Zhan Cui. The KRAFT architecture for knowledge fusion and transformation. *Knowl.-Based Syst*, 13(2-3):113–120, 2000.
10. Twikiguest. https://authdev.it.ohio-state.edu/twiki/bin/view/Main/VirtualOrganizations.
11. Gerhard Weiß, editor. *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. MIT Press, Cambridge, MA, USA, 1999.