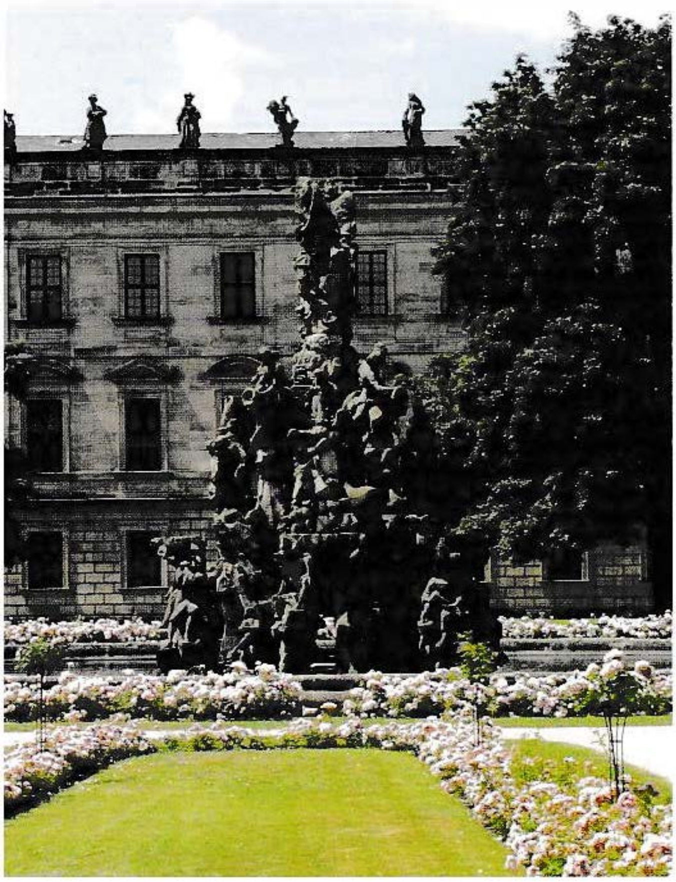
6th Workshop on Agent-Based Simulation

Penny Baillie-de Byl
Rainer Rimane
(Editors)

September 12-15, 2005
Erlangen
Germany



Workshop 2005

Friedrich-Alexander-Universität Erlangen-Nürnberg

ENTERPRISE DYNAMICS

Elitenetzwerk Bayern

SCS Publishing House e.V.

in cooperation with

ASIM – Arbeitsgemeinschaft Simulation
and
EUROSIM

ISBN 3-936150-44-3 ASIM

Printed: Gruner Druck GmbH, Sonnenstraße 23b, 91058 Erlangen

We thank our Sponsors

Friedrich-Alexander-Universität
Erlangen-Nürnberg

ENTERPRISE
DYNAMICS

DUALIS®

Elitenetzwerk
Bayern

ABS 2005

Content

Preface

The 6th Workshop on Agent Based Simulation is hosted by the University of Erlangen-Nuremberg in Germany. It should have been in Australia in April this year but for several reasons it could not be held as planned. So this workshop was postponed and transferred to Erlangen and collocated with the 18th Symposium of System Simulation held by the German Arbeitsgemeinschaft Simulation ASIM. Participants of both conferences can visit the lectures of the other conference as well. Our thanks go to Professor Ulrich Rüde the Chair of the ASIM Conference who gave us the chance to run the ABS6 here in Erlangen.

The 1st Agent Based Simulation Workshop was held in Passau by Bernd Schmidt and repeated twice at the University of Passau. Then the ABS-workshop went to Montpellier and last year to Lisbon.

A conference like this needs a team of dedicated individuals to prepare logistics and execute the necessary details to make the meeting possible. Therefore my special thanks go to Jan Willies. His contribution was invaluable. All the members of the international Program Committee worked hard to organize the review and select the best 10 Papers out of 16 submissions.

This proceeding shows the vitality of the multi-agent system modeling simulation community by the diversity of domain in which MAS is applied:

Special thanks to our keynote speakers Mikhail Prokopenko and Richard Zobel.

A scientific excursion is planned. SIEMENS Medical Solutions invited our participants to visit them.

We are grateful to our sponsors Incontrol Enterprise Dynamics GmbH and "Bavarian Graduate School in Computational Engineering", one of the new programs funded within the Elite Network of Bavaria.

I hope you will profit from the days here in Erlangen. I wish you all a fruitful and interesting conference with excellent scientists from all over the world.

Rainer Rimane

in the name of the ABS-workshop organization team

Keynote

Richard Zobel

Modelling and Simulation of Tsunami Waves for Improved Defence Provision in Phuket

Richard N. Zobel
Email: rzobel@phuket.psu.ac.th
r.zobel@ntlworld.com
Department of Computer Engineering
Faculty of Engineering
Prince of Songkla University, Phuket Campus
80 Vichit Songkram Road, Kathu
Phuket, 83210
Thailand

Abstract

Tsunami sea wave simulation is both complex and difficult, generally requiring large parallel computing facilities. There are two scenarios: the first, represented by compressible water Navier-Stokes equations representation, used for both deep sea waves and shallow sea water waves, and the second, non-compressible water Navier-Stokes equations representation, used mainly for shallow water coastal situations. The latter requires less computing power. Prediction of earthquakes and tsunami and provision of warning systems is not the object of this paper. Simulating the local effects of possible tsunami, is required to enable cost effective and efficient provision of local defences against future events and is the key concern discussed in this paper.

Keywords: Water wave modelling and simulation, tsunami, Navier-Stokes equations, 26 December 2004 disaster, future tsunami defences.

1 Introduction

The author was working in Phuket from three days after the tsunami for the following two and a half months, and was able to observe the local effects at the beaches around Phuket, the largest island in the Andaman Sea, off the South West coast of Thailand.

It is clear from still pictures and video taken at the time, that the situation at each location is unique and relates to both local factors and the magnitude, phase and direction of the sea wave in relation to the local geography of both the sea bed and the beach area. Some beaches were badly damaged and others less so, some were badly damaged at the ends of the beaches. Others suffered because the land behind the beach was very flat, and some areas were buffered by the mangrove forest.

The author considers the photographic and video evidence acquired at the time and later, in relation to research required to establish the key factors responsible for the damage, loss of life and injury for this specific tsunami. It continues by considering the possible effects for future tsunami with different characteristics, by using simulation to predict the consequences for the affected areas. Possible defences which could be provided to minimise the effects of future tsunami are discussed, with conclusions aimed at those who may need to consider these defences. The emphasis of the tutorial is on how modelling and simulation can be used to study the possible effects of future tsunami with respect to affordable provision of defences at the sites likely to be affected.

2 The Earthquake and Tsunami, 26 December 2004

The seaquake and ensuing tsunami, which caused so much damage and loss of life in and around the Indian Ocean, was one of largest events of its type for some time. In this paper, the author discusses the event and its effects in relation to the possibilities of using modelling and simulation to assist with understanding such phenomena and of attempting to minimise the effects on human coastal populations and property for the future. By a happy accident, the author arrived in Phuket just three days after the event and has been in a position to collect local information on the affected areas and of the sequence of events, especially the earthquake felt in Thailand before tbe tsunami arrived, and opinions from other parts of South Thailand, particularly from Hat Yai and Songkhla in the eastem part of the peninsula of Kra.

- **Plate Tectonics**

Plate tectonics is well researched and many publications exist on the subject, which attracts a wide audience ranging from the mildly curious to the serious scientific academic mind [1,2,3] The Indian Ocean has a complex of ocean ridges, mid-ocean areas where magma wells up from below the ocean flow, solidifies, and gives rise to ocean floor spreading in both directions from the ridge. In connection with the events of the December 26, 2004 earthquake and resulting tsunami, the ocean ridge concerned was the so-called 90-East Ridge, to be found in many atlas publications, in particular that illustrated in [4].

In the situation concerning the recent major event, the Australian-Indian plate is moving towards the Eurasian plate and has been in collision with it for millions of years. More specifically, the Indian plate has been pressing against the Burma (Myanmar) micro plate. The pressure has built up over a long period of time. How long is not known, but the pressure was eventually released on 26 December 2004 in a large movement, Richter scale 9.3, which saw the Indian Ocean floor move some 15m or more towards Indonesia. The Nicobar and Andaman Islands are on the Burma micro plate along with the Northern tip of Sumatra, which is Aceh. The interface, or collision, between these tectonic plates is called a *subduction zone*, which is where an ocean plate slides underneath a continental plate. This gives rise to island arcs, in this case those including Nias Island, and the major island cbains, such as those forming the Indonesian islands of Sumatra and Java and the Nicobar and Andaman islands. These islands often include many active volcanoes, and some of which are a volcanic island, as is the beautiful island of Bali.

- **The 90 East Ridge**

There are several ocean ridges in the Indian ocean. The 90 East Ridge, is an Indian ocean ridge, running in a North/South direction, roughly along the 90° East line of longitude, all the way from Australia to India, a distance of some 5000km. Associated with the subduction zone to the East of the ridge, are several deep trenches, notably the Java Trench, depth around 7km, and the Sumatra Trench, depth around 6km [5]. It is a very ancient structure and is the site where the Indian Ocean floor slides under the Burma Micro Plate.

3 The Tsunami

The seaquake occurred West of Aceh in Sumatra, Indonesia. This would seem to suggest a single epicentre or point of failure. However, this appeared to extend to a line of activity stretching perhaps 1200km, an extraordinary distance, which affected the propagation of the tsunami.

For the recent tsunami, it subsequently became clear that the total inverted rupture duration was about 200s and the peak slip was about 20m. The rupture propagated from the original epicentre towards the northwest for nearly 400km with a speed of 2.0km/s, along the subduction line east of the of the 90 East Ridge, producing a succession of earthquakes. This substantially changed the nature of the tsunami, especially in its effect on Phuket and the adjacent mainland areas of Phang Nga and Krabi. Following this development, part of the tsunami wave front became a North/South line travelling East towards Sri Lanka and West towards Phuket. This is distinctly different from that emanating from a single point epicentre.

- **The Tsunami in Phuket**

Phuket Island, and much of South and West Thailand suffered an earthquake at around 07.59 that morning, around one hour and forty six minutes before the tsunami arrived. In Hat Yai, on the Eastern side of the Kra Peninsula, some pieces of masonry fell from some tall buildings. The Earthquake was also felt in Bangkok and Chiang Mai, causing some damage and not a little concern by the residents.

The Tsunami arrived in Phuket and the nearby areas of Krabi and Phang Nga at approximately 09.45. The two waves destroyed many houses and businesses in just 10 minutes, in addition to killing and injuring many thousands of local people and tourists. The largest tourist beach in Phuket is Patong beach, along curved beach with beautiful white sand, gently shelving into the sea. This means that swimming is good and safe for children and non-swimmers. It is also bad for tsunami, because of the shallow water. Normally, the sea is calm and safe. After the tsunami, the beach area has been swept clean, and most of the beach buildings, sun beds and other fragile facilities, have been destroyed, as seen in figure 1.

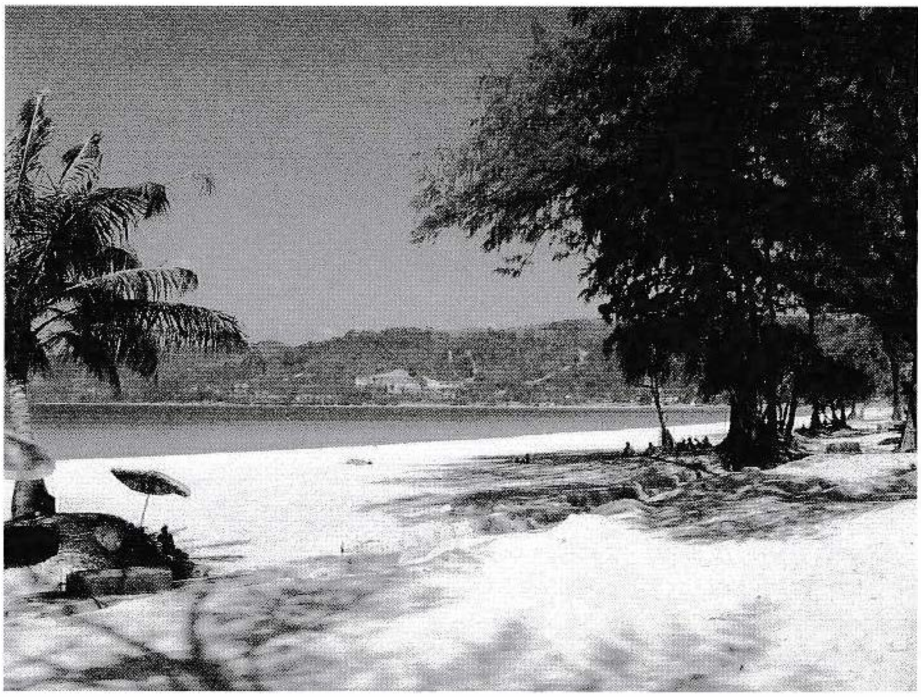


Figure 1: Damaged, almost deserted, Patong beach, North side, 23/01/2005

To the North of Patong beach, over the steep hill and down to the next beach, is Kamala beach, shown in figure 2. Kamala was almost wiped out, because it is so flat, all the way through the beach area, the village behind it and across the main road beyond. The figure shows a plot of land, which was perhaps a bar, but is now just a patch of tiled concrete, with its property boundaries defined. Some houses survived, others simply disappeared. Facing inland, in the opposite direction, is the remains of a single storey restaurant complex, substantially damaged, but was already under repair and reconstruction four weeks after the event.

Figure 2: Kamala beach, showing a destroyed building, now just a plot of land.

To the north, up and over the hill to the next beach is Surin Beach, another beach popular with tourists, but not usually so crowded as Patong Beach (Figure 3).

Figure 3: Steeply shelving Surin beach, with relatively little damage.

This is a steeply shelving beach where swimming is good in January and February, but dangerous later in the year when the normal sea waves exhibit a significant undertow, which drags swimmers out to sea to drown, because the swimmer is so quickly out of their depth. A consequence of the shelving is, however, that the tsunami wave was much smaller and consequently, most of the beach facilities and restaurants survived.

Figure 4: Nai Yang beach near the airport, showing a damaged long-tailed boat.

In the North West of Phuket, near to the International Airport, is the beautiful Nai Yang beach. Traditionally undeveloped, the beach facilities were destroyed but relatively quickly rebuilt. Illustrated in figure 4, is a badly damaged long-tailed boat. Nearby, several of these boats were being repaired in February, in addition to some new boats being constructed along with associated buildings and restaurants.

4 Modelling of Sea Waves

This subject is complex and concerns the general Navier-Stokes three-dimensional (3-D) partial differential equations for viscous, compressible flow as follows:

- **Eulerian Conservation of Mass:**

$$(\partial/\partial t + U \cdot \nabla)\,\rho = -\rho \nabla \cdot U$$

Where: U is the particle velocity, and ρ is the water density

$$\nabla = \partial/\partial x\,\mathbf{i} + \partial/\partial y\,\mathbf{j} + \partial/\partial z\,\mathbf{k}; \quad \text{where } \mathbf{i}, \mathbf{j}, \mathbf{k} \text{ are unit vectors}$$

and

$$U = U_x\,\mathbf{i} + U_y\,\mathbf{j} + U_z\,\mathbf{k}$$

Conservation of mass is of fundamental importance to most systems. In this context the mass balance must include all mass arriving in a given 3-D volume and all mass leaving that volume during a computational time period, in addition to the mass already present in that volume prior to that time period.

- Eulerian Conservation of Momentum:

$$\rho(\partial\partial t + U \cdot \nabla) U = -\nabla \cdot \sigma + \rho g$$

$$\sigma = \delta_{ij} P - S_{ij}$$

Where: P is the water pressure, g is the gravitational constant, and S_{ij} are the total viscosity deviators (e.g. $S_x = q_x - q$), where q is the viscosity).

Conservation of momentum is a second fundamental consideration in that the systems sum of the products of individual mass elements and their velocity must remain constant.

- **Eulerian Conservation of Energy:**

$$\rho(\partial\partial t + U \cdot \nabla) I = -\sigma : \nabla U + \lambda \nabla^2 T$$

where: $\sigma : \nabla U = \sigma_{ji} \partial U_i / \partial X_j$

I is the internal energy, and λ is a real viscosity coefficient.

Conservation of energy is a little more tricky. However, it must be remembered that energy is only transferred from one form to another and is consequently not lost.

This set of equations may be simplified by assuming the flow is incompressible ($\rho = \rho_0$)

The equations may be further simplified by excluding viscosity. Viscosity terms may be included if the fluid is incompressible, with or without shearing forces.

Further details and a full discussion of these equations may found in Mader [6].

In computational terms, each of these equation sets must be converted into finite difference equations before solutions may be computed digitally. In order to do this, assumptions must be made in relation to the most appropriate form for the representation of differential terms to achieve an acceptable accuracy of result. To represent such a large volume of water, it is necessary to carefully consider the spatial resolution, which needs to be adequate to represent the spatial frequencies, or wavelengths of the expected sea waves, but also needs to minimise the computational cost. This is always an awkward compromise. The temporal aspects must also be considered in terms of minimum acceptable sample rate in this respect. It is important to appreciate that for these systems, such as water wave systems, the consequences of having relatively low damping needs to be considered in relation to the convergence of the solutions. In relation to the shallow water simulations, these aspects also need to be carefully looked at, since in shallow water the wavelength becomes considerably shorter and the wave height considerably larger. In addition the volume under consideration becomes much smaller.

5 Possible Defences against Tsunami

Tsunami arise because of earthquakes, landslides, cavity collapses and other related sudden major movements of land above or below sea or water surface level. A few can be foreseen and measures possibly taken to prevent them. An example of this is the Lituya Bay mega tsunami [7]. This type of landslide, perhaps on a smaller scale, might be recognised as a hazard and steps taken to prevent it happening before it occurs. Several mudslides caused by removal of trees also fall into this category.

However, major earthquakes cannot currently be accurately predicted, and as yet, it is not possible to determine if a particular earthquake produces a tsunami. Hence it must be accepted that tsunami will occur in the future and could give rise to major loss of life, injury and financial loss. Tsunami warning systems exist, or are planned for the near future, for possible affected areas, especially for the Indian Ocean. However, it seems that a number of possibilities exist to take affordable measures to limit their effect.

- **Defence Possibilities**

It was observed that in Krabi, villages which lay behind mangrove forests were far less affected that those here the forest had been removed to make way for fish farms or other commercial activities. Perhaps minimisation of the scale of mangrove forest removal could also minimise the damage caused by future tsunami. Alternatively, man-made structures might be created which in some way mimic nature in this respect.

Figure 5: Mangrove forest Phang Nga Bay

Another possibility is for refuges to be constructed to provide escape to an upper story when stranded in a flat area, such as Khao Lak in Phang Nga Province to the North of Phuket Island and also for the centre of beach areas in Phuket, such as Patong and Kamala. Such structures would have on the first (ground level) floor only stairs and supports for the floors above, allowing the sea to flow through it. Changes in the shallow seas close to the beeches might also be effective in some circumstances, perhaps by dredging, or of dumping sand or rocks. Modification to the built environment might also help if the possibilities of tsunami attack are considered at the planning stages. Changes to the vehicle parking policy and provision could also help here.

However, none of these can be seriously considered until predictions of possible tsunami sizes, types, directions, etc, can be made available by appropriate modelling and simulation. When this is achieved, the efficacy of the measures described above can also be assessed and costed. All of this requires computer simulation to obtain typical possible tsunami data for the areas under threat. Further, simulation of the effects of such tsunami simulations on building structures and the built environment also require computer simulation, as does the parking arrangements.

6 Modelling and Simulation of Possible Tsunami in the Indian Ocean

The source of the tsunami in the northern part of the Indian Ocean, which would affect the western coasts of South Thailand, is most likely to be the subduction zone earthquakes of significant magnitude, associated with the fault line which includes the Sunda Trench, commencing from the centre of the west coast of Sumatra, right up to the Andaman and Nicobar islands in the north. Both single and multiple earthquakes (as for the recent events) need to be considered. Simulation of the effects of such earthquakes on possible tsunami generation is needed, but not easy to model.

Hydrological data is required to model the seafloor boundaries. However, the recent tsunami has caused changes to the sea floor, requiring new hydrological surveys to be carried out, especially in the Straits of Malacca an import sea route, and this aspect is also particularly relevant to the West coast of Thailand. The possible effects of reflections from islands and undersea rock formations also need to be considered. Such studies will provide samples of input data for shallow water studies in the coastal areas. Full Navier-Stokes compressible water wave simulations are required at least to validate any simpler models used to reduce the costs of simulation. This is true for both deep water and shallow water studies.

The tsunami simulated for the Indian Ocean, and particularly for the Andaman Sea, are required as input to the shallow water simulation studies for the general and specific bays and beach areas of Phuket, Phang Nga and Krabi. There is, of course, the problem of change of scale and consequent spatial resolution, resolved perhaps by interpolation. However, there are also a number of potential problems in achieving a reasonably accurate representation of the resulting sea wave motion.

- **Tsunami Modelling Problems**

First, it is necessary to obtain recent or current data of the sea floor hydrological data, in relation to the type of deposits and exposed rock formations as evidenced for example, in the video footage showing the sea receded to the horizon at Patong beach, prior to the arrival of the tsunami wave.

Second, it is important to obtain estimates of the sand and other materials picked up by the first wave. This affects the local water density, and also the way in which such mixtures of solid and liquid matter behave probably needs to be studied.

Consequent upon the first tsunami sea wave arriving, debris will be collected from the beach area and carried out into the bay. Some of this will be floating wreckage, including sun beds, and umbrellas, other material will be heavier and sink, but will still be carried away by the force of the water. Some of this debris and the suspended sand will subsequently be deposited in the bay area, changing the sea bed data, which in turn will require changes to the model for the second sea wave. It is not clear how serious this is for the modelling process. However, it is clear that some small islands, sand bars, sand spits and sand banks, together with some beach features, were modified by the recent tsunami. Indeed some small islands have disappeared; some new islands have appeared, and some beaches have been modified. This was very apparent from the air during the author's several flights over the area with local Thai colleagues after the event (Figure 6). It was also clear that very significant disturbance of the sea bed in the coastal areas had occurred, as viewed from the aircraft when the author flew from South to North on the Western side of Phuket island when arriving from Bangkok on 29 December, just three days after the tsunami.

Figure 6: Small island between Phuket and Krabi after the tsunami

7 Application of Simulation Results to Affordable Defences

The result of the simulation of possible tsunami scenarios at specific beaches requires interfaces with other simulations. Specifically, the effects of such tsunami need to be applied to a variety of building structures with a view to determining their susceptibility to such sea waves. In this respect, it is interesting to observe, that a two storey restaurant, close to the beach at Patong, had its ground floor cleared, leaving just the supporting pillars holding up the restaurant above, which was apparently undamaged. Resulting from these observations, it is relevant to propose that some existing designs and proposals for designs of new buildings be considered to replace those destroyed. Specifically, it is already obvious that basement areas in department stores, whether for car parking, shopping or other human activities, should be avoided where these are in areas liable to flooding from tsunami attacks. Further, studies of the tsunami on a variety of designs for the built environment in the vulnerable areas close the beaches can be made to determine favourable designs for reconstruction.

Perhaps of greater significance, is the provision of refuge buildings, specifically designed to allow escape to higher level, above the invading sea. This is especially important in the flat areas of Phang Nga, such as Khao Lak and the central area of Patong beach. The parking of vehicles, whether for delivery in connection with business or for tourists visiting the beach on vacation is a problem everywhere, but especially difficult in Phuket, and getting worse. The consequences of a tsunami picking up vehicles and smashing them into shops and restaurants full of staff and customers is clearly shown in photographs taken at the time of the disaster in Phuket. A possible solution is to construct multi-storey car parks supported on columns, allowing the sea to pass safely underneath. This would ease the parking problem, reduce the danger and also provide refuge for people escaping from the arriving tsunami. Of course, such car parks must NOT have underground levels, unless they are clearly above possible inundation. Such buildings with supporting columns are common in Thailand, providing storage and vehicle parking underneath and reducing the risk of entry of snakes and other undesirable visitors to domestic apartments as may be seen in the accommodation blocks at the Phuket Campus of the Prince of Songkla University, up in the hills of Kathu district, as illustrated in Figure 7.

Figure 7: Accommodation block, Phuket campus

Many other aspects of where modelling and simulation might be useful must also be considered. For example, the design of the cross section of the pillar supports needs to be optimised to provide minimum resistance to sea waves, in order to maximise the possibility that upper storeys survive a tsunami attack. Examination of video footage of the tsunami at Patong Beach shows that, after the tsunami waves had initially receded, the sea took on a violent chaotic appearance and that this was replaced by a large whirlpool. Perhaps these damaging activities could have been ameliorated by modifying the sea floor close to the beach by dredging or by dumping sand or rock to damp out this motion. Finally, the possibilities of creating moles out to sea to reduce access of large sea waves to the beach areas, as can be seen at Schevningen, in the Netherlands and at other places.

8 Conclusions

The tragic events of 26 December 2004 in the countries bordering the Indian Ocean have lead to a desirable initial response of creating a tsunami warning system to enable coastal areas likely to be affected to be evacuated. However, the flat, low lying areas of Khao Lak in Phang Nga, Thailand and Aceh, in Sumatra, Indonesia provide little in the way of possible escape to a higher level above the invading sea. This paper considers the possibilities of the use of computer modelling and simulation of sea waves to determine the feasibility of making economic defence provision for affected areas in Thailand, which may also be relevant tin other countries.

The paper outlines the stages of the work required, and highlights some of the likely major technical difficulties in carrying out such a novel study. Perhaps the most important of these difficulties lies in the problems of interfacing the deep sea model to the shallow sea model and in turn to the models required to determine the effects of the tsunami on building structures and the built environment.

The problems of determining the details and modelling of the effects of sand and debris following the arrival of the first tsunami wave are mentioned and seen to be formidable.

Finally, possible economic solutions are proposed in relation to more robust and safer building practices for the affected areas and to the contribution of purpose built or dual purpose refuge buildings for escape from such damaging sea waves.

9 Bibliography

[1] Moorland School. 2005, Clitheroe, Lancashire, BB7 2AJ England. *Plate Tectonics*. Earth Science for Schools. www.moorlandschool.co.uk

[2] Keary, P. & Vine, F.J. 1991. Global Tectonics. *Blackwell Scientific Publications*, 302pp..(2nd Edition)

[3] Press, F. & Siever, R. 1998. Understanding Earth (2nd Edition) W.H. Freeman, p.504–535 + CD-rom

[4] Insight World Travel Atlas, Verlag Wolfgang Kunth GmbH, Munich, Germany. p 315. ISBN 981-258-070-0

[5] Weis, D., Frey, F.A., Saunders, A. and Gibson, I., 1991 Ninety East Ridge (Indian Ocean): A 5000 km record of a DUPAL mantle plume, *Geology, 19*, 99-102.

[6] Mader C.L. 2004. Numerical Modeling of Water Waves. 2nd ed. CRC Press. ISBN 0-8493-2311-8. pp 1-27

[7] Mader C.L. 2004. Numerical Modeling of Water Waves. 2nd ed. CRC Press. ISBN 0-8493-2311-8. pp 203-214

10 About the Author

RICHARD ZOBEL graduated in Electrical Engineering from London University in 1963. His first experience of simulation was obtained during 1962-66 at Sperry Gyroscope whilst working on naval surface to air missiles, using mainly valve analog computers. His Ph.D., obtained in 1970 at Manchester University, concerned hybrid analog-digital computing. As Lecturer and Senior Lecturer he became involved in digital signal processing, instrumentation and design environments with special emphasis on the simulation aspects of real-time embedded systems. He is a Committee Member and former Chairman of the United Kingdom Simulation Society (UKSim), Former Secretary of the European Federation of Simulation Societies (EUROSIM), and was a European Director of SCSI. He was presented with the Outstanding Service Award of SCSI, the Society for Modelling and Simulation International, San Diego, California, USA, at the Dresden Conference in 2003. His current research interests concern distributed simulation for non-military applications, model re-use, distributed simulation model databases, issues of verification and validation of re-useable simulation models and security for distributed simulation under commercial network protocols. He is now retired, but still very active. He is currently teaching Computer Engineering during the winter months at the Prince of Songkla University, Phuket and Hat Yai Campuses in South Thailand.

Papers

On validation of individual-based models

Jochen Wittmann

wittmann@informatik.uni-hamburg.de
Universität Hamburg, Fachbereich Informatik, Arbeitsbereich Technische Informatiksysteme
Vogt-Kölln-Straße 30, 22527 Hamburg, Germany

KEYWORDS
Model validation, individual-based models, experimental design

ABSTRACT
This paper analyses typical experimental set-ups for individual-based models on a not-aggregated level of model description in comparison to conventionally aggregated models.
It postulates that for real-world-applications additional assumptions become necessary which concern to the type and the parameters of the data transformation between the aggregated and the non-aggregated level.
The structure of the problem is analysed and typical scenarios for model usage and validation are listed. General methodological deliberations for each of these scenarios are made which offer a guideline for correct experimental design in order to validate the corresponding models.

1. Introduction

The object oriented modeling paradigma has established during the last years and leads especially for the application areas biology, ecology to its specialization in the form of so called „individual-based" models. This paper will not go into further discussions on the definitions of "individual-based" in contrary to "individual-oriented" or even "agent-based" modelling. A comprehensive summary concerning this topic can be found in (Ortmann 1999). However, the paper will analyse the validation step during a simulation study if an individual-based approach has been chosen.

With regard to the main application areas of individual-based models, which mainly are applied in domains without exact physically derived model descriptions, this important phase in a simulation study attracts special attention.

It is typical for individual-based models to model the reality, the objects under observation, and their behaviour in a very natural way by close analogy between the real world objects and the objects – or individuals – used on model description level. Therefore, this modelling paradigm leads to a class of models which satisfy the criteria of adequate and easy understandable model structure on a very high level.

On the other hand, these models often are associated with the disadvantages caused by their demands concerning processor time and memory. This problem is a direct consequence of the non-aggregated model description and seems to be the price the user has to pay for comprehensibility and transparency on model specification level. This problem, too, shall not be discussed here.

This paper will focus on a further problem field which seems to be neglected in the main discussion of individual-based models: the problem of model validation. The pretended simpleness in model description often implies the need for a highly sophisticated analyse of the model and its results in the phase after the runs, in validation and interpretation of the results. In this situation, this paper analyses typical dilemmas and tries to give hints for a proper determination of the range of validity for individual-based models.

2. Simulation on local and on global level

To understand the problems concerning validation, we start with a view on the general design of a modelling and simulation study based on the individual-based paradigma. Figure 1 depicts the course of the argumentation in comparison to the use of a "conventional", i.e. non-individual-based model.

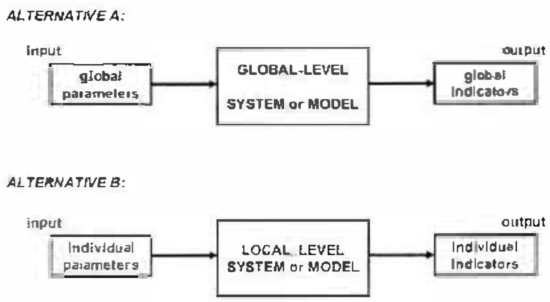


Fig. 1: Comparison between individual-based and non-individual-based modelling studies

Both alternatives work according to the same basic scheme: Alternative A shows the situation for a conventional model on global, which means here accumulated, level in model specification. The modellet and experimentator is interested in the effects of a change in a global parameter. This parameter is set for the simulation and after the run an other parameter on global level, a global indicator variable is observed.

Example: global input parameter is the reproduction rate of a population, the model is a common differential equation model for the population dynamics, and the model result is the population for a future point in time. Input, output, and model equations work on highly aggregated data for the population, which mirror the situation on individual level in statistical sense.

On the other hand, alternative B describes the system dynamics on the individual level. Example: For the population dynamics, a possible input parameter would be the mean number of children a woman gets during her life, one would have to model the interactions of the individuals and would be able to derive an individual curriculum vitae for each of the individuals. At the end, the actual number of children each individual has got would be the observation parameter on this level.

Both alternatives are proper implementations of the same basic modelling and simulation approach. The experiment deals with the objects input variable, the model itself, and the output. Accordingly the three basic tasks are identified: system identification (input and output given), forecast (input and system model given), and control (system model and output given). Differences between the alternatives A and B can only be found on the level of model description: In the first case, the complete model is specified using the population number as a cumulated value. The second case specifies the behaviour of the individuals and produces the population number as a dependent variable of the set of interacting individuals. Naturally, both model approaches have to be parameterised and validated on their specific level of model description. In consequence, even the results can only be interpreted and exploited on the level of specification the model offers.

3. Data transformation between the levels

There is one observation which appears from the simple description of the experimental set-up described so far: During the simulation run an individual-based model produces the curriculum vitae of the set of individuals under observation. If the experimenter is interested in more general model quantities, a recalculation and evaluation of those raw data will be necessary. (In our very simple example this recalculation step is realized by a simple summation of the individuals living at a certain point of time and could be realised as a dependent model quality as well.) This argumentation implies a change of modelling level for data evaluation and interpretation (i.e. from level A to level B) concerning the two alternative scenarios introduced in figure 1.

Similar and much more complicated transformations from one level to the other can be necessary in a number of simulation experiments which deal with individual-based models.

In general, the change of levels is usefully applied if missing information on the one level is replaced by or can be derived from well known information on the other level. Such a level change can be done on the input-side as well as on the side of the outputs.

So far there are no problems in the experimental set-up and the situation can be recapitulated graphically by figure 2. The difficulties, however, arise when the model has to be validated and the situation escalates if there is a lack of comprehensive system data.

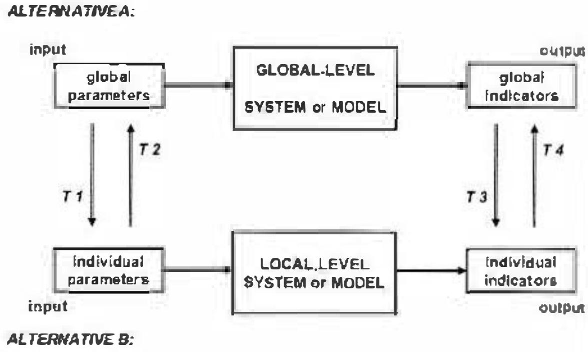


Fig.2: possible transformations between the levels of model description during experimentation

Example: For the very simple population dynamics example the transformations T1 to T4 introduced by the figure shall be exemplified:

1. A known mean life expectancy is transformed into determined ages for set of identical model individuals.
2. In a statistical sample size and weight of persons are measured, the mean values are used as parameters for a model on global level.
3. A certain mean value for energy consumption of a region has to be allocated to individual energy consumption values for each individual living in the region.
4. The total population number is summarised by counting the model individuals at a certain point of time.

Usually, the transformations from the individual level to the global level are evident and easily to execute. In this direction, there exist data on detail level which have to be aggregated to a more general, often statistical parameter value on the global level.

Transformations in the other direction are not possible without at least two further assumptions:

1. the type of distribution of the parameter transformed (e.g. uniform, normal, ...)
2. parameters of the distribution, such as mean value, variance, ...

But even the very simple transformation of type 4 (individual level to global level) can be more than a simple summation and has to be considered with carefulness. An example: The individually directed voices during an election could be weighted. Therefore an additional set of weight parameters has to be specified for the model and the corresponding aggregation function has to be calculated for a correctly executed level change.

4. The problem

The argumentation so far explains the theoretical design of simulation experiments on the both levels introduced. However, in praxis and especially in the praxis of the application domains which like to use individual based models of cause of their easy and structure adequate model description facilities, the missing data forces to a more sophisticated, combined experiment design crossing the levels. Therefore transformations become necessary which imply additional parameters.

The systematic problem of these parameters is that their values cannot be acquired separately. If it would be possible to do so, the transformation and the level change would not have been necessary.

On the other hand, proper parameter identification needs measurements on both levels to identify the transformation parameters first and to calculate their values afterwards. This is an inherent contradiction of the experimental design. It is caused by the situation of system data and will not be dissolved by additional data acquisition in the real system.

For the modelling and simulation study follows: A separate validation of the assumptions concerning transformation parameters and their values is not possible. They have to be an additional task within the global model validation process.

To formulate constructively: The model experiments have to be designed in a manner that

1. the model results are independent of these transformation parameters, or
2. there is a proper distinction between the influence and effects of the transformations and their parameters and the effects of a change in the model parameters which in fact are under observation to achieve the experiments objectives.

In both cases the validation implies additional restrictions for the experimental design. The experiments have to assure that a statistical distinction between the effects of the transformations and those of the intended classical investigation according to the tasks identification, forecast, and control becomes possible.

Naturally, this problem escalates because even in the model there are variations in parameters to test, which are caused by uncertainties concerning model parameter values and even model structure. Figure 3 concludes these possibilities in argumentation for the different alternatives in experimental design.

It is obvious that the additional parameters make the study much more complex and the intended direct causality between the experimental parameters and their effects becomes more and more difficult to extract.

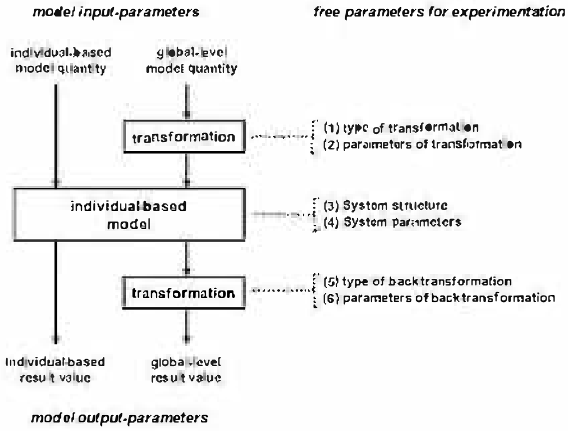


Fig.3: data-flow and free experimental parameters

5. Possible experimental designs for validation

So far, the need for sophisticated statistical methods for validation has been elaborated. Furthermore it is obvious that it will not be possible to validate the additional parameters separately, because there are no (or at least: not enough) system data.

In this situation, four possible and typical experimental designs shall be analysed with regard on a feasible model validation. The objective is to demonstrate the general argumentation and to explain the logical consequences of the initially chosen experimental design.

5.1 Individual behaviour under observation

The most obvious motivation for building individual-oriented models is to investigate in the behaviour of tbe individuals itself. This is represented by alternative B from figurc 1. The experimental design is without any modification as it is usual in modelling and simulation because all operations take place on individual levcl. For the validation system data and model data have to be compared and the range of validity has to be determined from these deliberations. Concerning the structure of system and model equal assertions can be made, and the free parameters (numbers (1), (2), (5) and (6)) from figure 3 are not relevant in this case.

However, one sbould pay attention to tbe format of simulation results: To be accurate, only the life data for individuals are observed on the individual level. There is no aggregation of the data at all. Any aggregation would be interpreted as a cbange to the global level and would imply the necessity of a transformation of type T4 with tbe corrcsponding parameters and difficulties.

These deliberations lead to the next experimental scenario:

5.2.1 Structural adequate models for global processes

The motivation for this design variant comes from model description methodology: There exists the presumption that a model code as well as a program code is easier to understand and more efficiently to maintain if its structure mirrors the real world structure of the modelled systcm. With this background, the individual-based model description seems to offer the optimal level of comprehensibility because tbis model specification paradigm propagates to be nearly completely adequate.

For the validation context, one interesting observation must be made in connection with this approach: Even thougb the interests of the experimentation lic on the global level, model description and simulation work witb the non-aggregated level. Therefore the model holds a scale in detail, wbich is not necessary for tbc level of results the experimentation intends. If the information on the detail level can be provided, this approach is very self-explaining and the advantages of the evident model structure overweight the demands in run-time those models usually need.

If there is a lack of information concerning parameters on the individual level, there are lots of additional bypothesis concerning type and parameter values of the transformations to calculate and validate, a task which has to be solved by data collected on the aggregated level solely. Thus, a serious validation for tbis kind of models succeeds only with great efforts in statistical determination of the missing parameters. In praxis the modeller will have to weight wbether the adequate model structure will be worth tbese investments in statistical procedures. These deliberations show that the evaluation of tbis experimental design scenario has to be made for eacb application separately. The balance betwecn investments and effort as described above should be considered very carefully.

5.3 Measurements are not possible on the desired level of model description

This scenario is very similar to the preceding one; however, in this case the experimenter has no choice between the alternatives in level because a missing access to the data on tbe one level forces bim/her to substitutc the missing information by investigations on the other one.

To be able to parametetise, validate, and work with the model at all, at least one of tbe transformations has to be specified and parameterised. Here the efforts are the prize for capacity to act not only the prize for an adequate, a nice model structure.

The limitations concerning accuracy and validity of the model have to be accepted. The experimental design bas to be very sophisticated but the way of additional transformations is tbe only one, which provides access to a region of knowledge otherwise completely inaccessible.

5.4 Investigations on emergent behaviour

Highly interesting is an application field for individual based models not yet mentioned in this paper so far: the so called "emergent behaviour". In short, tbis means a bebaviour of a group or mathematically spoken a set of identical individuals which is observed when these individuals interact, communicate, and cooperate but which is no specificd explicitly within tbe bebaviour specification of the single individual (e.g. tbe organisation of tbe ants, swarms, ..).

It is evident that the use of individual based models is inevitable in this case. Here, tbe experiment focuses on one of our transformations: The purpose of the model is to describe individual bebaviour locally, let the individuals interact, and to observe behaviour of the group of individuals wbich has not been specified explicitly on the local level. The change of level is the trick: input on local, measurement of output on global level.

A further analysis touches the assumption that has bcen the base for all tbe deliberations before: the existence of well-defmed rules for aggregation.

This assumption is challenged by the assumption of emergent behaviour.

There is no transformation specification in the form of rules or functions! In contrary, the observations on global level are generated by the behaviour specification on local level exclusively.

So far the theory. In real world applications the investigations on emergent behaviour naturally are superposed by the problems in getting proper system data. Therefore, very often level transformations are necessary to avoid data lacks. These transformations have to be parameterised and validated as described before. To prove real evident behaviour properly it is inevitable to separate the transformation and its effects from the observations and investigations made to prove the emergent behaviour.

If the parameters of the transformation are not known, complex additional experiments are necessary to determine their effects first, and let the argumentation turn to the phenomena of emergent processes only if there are no more doubts concerning "technical" transformation parameters. Especially for validation these interacting effects have to be differentiated and isolated to make real causalities between local behaviour specification and global level parameters evident.

6. Concluding example

The well known predator-prey model shall serve as a very simple example to illustrate the problems and the argumentation for the different experimental set-ups.

Alternative A implements the model by the well-known set of two differential equations for the two populations. Alternative B specifies the same situation in an individual-based manner. The question has to be discussed, how information on the one level can be completed by data on the other level and how far the two levels provide support for validation.

First the (well known) suppositions for the differential equation model explicitly in advance:

1. The equations are valid only for large population numbers N.
2. The parameter values are based on equal distribution of the individuals on the field. (e.g. for the meeting probability)

To demonstrate the dilemma comparing individual-based and global model to each other, the following deliberations will be enough:

1. If the individual-based model is run with low population number N, there is a direct contradiction to the assumption 1 for the global model.
2. If the individual-based model is run with large population number N, there will be a contradiction with the assumption 2: If there are lots of individuals, the distribution over the area under observation will not be equal. Normally, there are groups of hunting predators with no prey in between them in one block and in an other region other groups of prey with no predators in between.

The consequence for the experimenter is now: Is the group building process just a mistake in model description or should it be interpreted as emergent behaviour? Often the answer of this question draws upon the data produced by the model on the other level. As explained, such an argumentation breaks the assumptions. There is no other way out than to specify the transformations between the levels, determine their parameters and validate the hypothesis on this statistically detailed level.

Concerning the validation of models by a second model of the same system but on an other level of detail the conflict is obvious as well: The change of model specification level does not replace detailed validation based on additional experiments with the model and normally even with the real world system.

7. Resume

The paper tries to give a structure to discuss the validation of individual based models by mentioning the separate data transformation steps within the global and the local modelling level and between the levels themselves.

It emphasises that each transformation has additional parameters for its own that normally have to be determined by additional statistical experiments.

A comparison of results gained by models on the different levels may be interesting, however, its statistical value for validation and interpretation of possibly appearing effects is negligible.

The proposed scheme does not provide an algorithm to solve the problems in using individual-based models but it tries to make the typical structures of argumentation using such models transparent and tries to give a guideline for the discussion of critical aspects and common problems using these types of models.

References

Ortmann, Jörg 1999. „Ein allgemeiner individuenorientierter Ansatz zur Modellierung von Populationsdynamiken in Ökosystemen unter Einbeziehung der Mikro- und Makroebene"; Dissertation am Fachbereich Informatik, Universität Rostock, 1999

TEAMWORK APPROACH FOR MODELING AND SIMULATION OF DDOS ATTACKS IN INTERNET

Igor Kotenko
Alexander Ulanov
St. Petersburg Institute for Informatics and Automation
39, 14[th] Liniya, St. Petersburg, 199178, Russia
{ivkote, ulanov}@iias.spb.su

KEYWORDS
Multi-agent systems, Agent-based Modeling and Simulation, Distributed Denial of Service (DDoS) Attacks, Network Security.

ABSTRACT

The paper considers an approach to modeling and simulation of Distributed Denial of Service (DDoS) attacks fulfilled by a group of malefactors. The approach is based on combination of "joint intentions" and "common plans" theories as well as state machines. The formal framework for modeling and simulation of DDoS) attacks is presented. The architecture and user interfaces of the Attack Simulator software prototype implemented and its evaluation results are depicted. The simulation-based exploration of the Attack Simulator prototype demonstrated its efficacy for accomplishing various DDoS attack scenarios. The framework and software prototype developed can be used for conducting experiments for evaluating computer network security and analyzing efficiency of security policy.

1. INTRODUCTION

Vulnerabilities of computer systems, permanently magnifying complexity of cyber-attacks and gravity of their consequences highlight urgent necessity for new approaches to information assurance and survivability of computer systems. One of the most harmful classes of attacks aiming at destruction of network resources availability is "Denial of Service" (DoS) (Mirkovic et al. 2002; Mirkovic et al. 2004). The purpose of DoS is isolation of a victim host. As a result of this attack the legitimate users can not access necessary network resources. Most of operating systems (OS), routers and network components are prone to DoS attacks that are hard to prevent.

The new type of attack arrived in the beginning thic century. It is called "Destributed Denial Of Service" (DDoS). To perform DDoS attacks malefactor needs to hack a set of computers ("zombies") at first and to run on them DoS programs to attack next targets. This makes hard to detect DDoS attack and to defense from it. The DDoS domain is becoming more and more complex. We observe now the great variety of different DDoS attacks and the continuous appearance of new types that break the defense.

The seriousness of the DDoS problem and the increased frequency of DDoS attacks have led to the development of numerous DDoS defensive mechanisms. Unfortunately, the existing theoretical basis that should support implementation of defensive mechanisms against such class of att poor.

According to our opinion, among many reasons, the a stipulated by weakness of fundamental researc consider defense against DDoS attacks as a t adversarial competition between security system malefactors' attacking systems, in particular, the n intending development of an adequate formal framew exploratory modeling and respective software archi for simulation of DDoS attacks and distributed de software components of computer network (Kotenk 2003).

Modeling and simulation of DDoS attacks and perf their analysis are very important for discovering co systems prone to DDoS, formulating d recommendations and developing effective pro methods.

The paper considers an approach to agent-based mo and simulation of DDoS attacks fulfilled by a gr malefactors. The goals of the paper are developm agent-based formal framework for specification of attacks and implementation of a software tool ma possible to simulate DDoS attacks.

The rest of the paper is structured as follows. *Sec* outlines suggested common approach for modeli simulation of DDoS attacks by imitating malef teamwork. *Section 3* describes the ontology of DDoS and specifications of structure and common sche operation of agents. *Section 4* determines architectu main user interfaces of the DDoS Attack si elaborated and its evaluation issues. *Conclusion* outli results of the paper.

2. DDOS COMPONENTS AS INTELLIGENT AGENTS. TEAMWORK-BASED FRAMEWORK FOR MODELING AND SIMULATION OF ATTA

By the analysis of present DDoS attacks it is poss reveal the division of DDoS software components b roles. At first, there is a "master" program which gath initial information about hosts in the Internet and obta access to their resources for starting "daemons" pro "Daemons" are the attack executors. They usually pr full access to compromised host for "master". "M coordinates "daemons" actions: it can exchange m with "daemons" and install on captured hosts new pr for further "daemons" propagation. "Daemon" rep "master" about its state. As soon as the DDoS net required size, "master" sends the messages about atta [illegible] target (or simply an attack signal) to all [illegible].

[illegible] of DDoS attacks we can see that each [illegible] program ("master" or "daemon") is an autonomous [illegible]ponent which has initial knowledge, can get [illegible] data from environment, has target and list of [illegible] to reach this target, and can interact with other [illegible]. These properties are peculiar to intelligent [illegible]. All DDoS components form a team of agents as they [illegible] operations for reaching the common long-time [illegible] of service attack) in a dynamic external [illegible] (the Internet) at presence of noise and [illegible] of opponents (components of security [illegible].

[illegible] research on teamwork is an area of steadfast [illegible] in multi-agent systems (Fan and Yen 2004). A set of [illegible] to formalization and simulation of the agents' [illegible] is known. For the organization of teamwork of [illegible] agents, we have used the base ideas stated in works [illegible] intention theory, the shared plans theory and the [illegible] theories of agents' teamwork.

[illegible] *intentions theory* (Cohen and Levesque 1991) the [illegible] which the agents act, are assumed to consist of [illegible] that can be associated with specific agent. [illegible] in which the agent is convinced are called its [illegible] of agent that are considered by it as the most [illegible] are called its goals. The mutual agents' beliefs are [illegible] beliefs of agent group. Agent team is said to [illegible] joint intention to complete an action if and only if [illegible] members have joint persistent goal to complete this [illegible].

[illegible] *shared plans theory* (Grosz and Kraus 1996) the [illegible] is believed to be the plan of joint fulfillment of [illegible] of actions by the group of agents. The main [illegible] of shared plan are as follows: (1) the group plan [illegible] the group (team) of agents should reach the consent [illegible] the instructions, to which they will follow in group [illegible]; (2) the agents should take up the obligations not [illegible] personal operations, but also on operations of the [illegible] as a whole (personal intentions how to make [illegible]; (3) each agent should take up the obligations on [illegible] of other agents (approved intention); (4) the plan [illegible] activity can have as components the plans of the [illegible] agents for the assigned operations, as well as plans [illegible].

[illegible] *theories* (Jennings, N. 1995; Tambe 1997; [illegible] and Pynadath 2001) the notion of joint persistent [illegible] used for building the scheme of agents' actions [illegible] and agents' communication protocols. The [illegible] of shared commitments is the basis for [illegible] and monitoring of team activity. This notion [illegible] while checking the state of goal and corresponding [illegible] the goal is achieved, not achieved, cannot be [illegible] or irrelevant in view of breaking conditions). The [illegible] joint intentions is used to describe the agents' team [illegible] of particular operators. The main goal in [illegible] is to provide the activity of agents according to a [illegible] scenario where each agent knows its place. The [illegible] theory supports required methods for solving [illegible] in form of shared plan (full or partial). This [illegible] specify the activity of whole team, agents' groups [illegible] agents and also the constraints determining agents' collaboration and communication. The joint intentions theory is used for structuring of shared plan, scenario of its execution and communication.

Considering the "master" and "daemons" specifics the most suitable approach to use is the combined theory. There must be a shared plan, because of need to provide agents work according to mentioned DDoS attack steps. Agents have a joint goal – to perform DDoS attack. However, "master" and "daemons" act each in one's own way. Individual actions and communications of agents will also be a part of shared plan.

The common (group, individual) intention and commitment are associated with each node of a general hierarchical plan. These intention and commitment manage execution of a general plan, providing necessary flexibility. During functioning each agent should possess the group beliefs concerning other team-mates. For achievement of the common beliefs at formation and disbandment of the common intentions agents should communicate. All agents' communications are managed by means of common commitments built in the common intentions. Besides it is supposed, that agents communicate only when there can be an inconsistency of their actions. It is important for reaction to unexpected changes of environment, maintenance of redistribution of roles of the agents failed or unable to execute some part of a general plan, and also at occurrence not planned actions (Tambe 1997).

The suggested technology for creation of the malefactors-agents' team (that is fair for other subject domains) consists in realization of the following chain of stages (Kotenko et al. 2003): (1) formation of the subject domain ontology; (2) determination of the agents' team structure; (3) definition of agent interaction-and-coordination mechanisms (including roles and scenarios of an agents' roles exchange); (4) specifications of the agents' actions plans (generation of attacks); (5) assignment of roles and allocation of plans between the agents; (6) state-machine based realization of the teamwork.

Formation of the subject domain ontology is an initial stage of the agents' team creation. Modeling in any subject domain assumes development of its conceptual model, i.e. set of basic concepts of a subject domain, relations between the concepts, and also data and algorithms interpreting these concepts and relations.

The agents' team structure is described in terms of a hierarchy of group and individual roles. Leaves of the hierarchy correspond to roles of individual agents, but intermediate nodes - to group roles.

The plan hierarchy specification is carried out for each role. For group plans it is necessary to express joint activity obviously. The following elements are described for each plan: (a) entry conditions when the plan is offered for execution; (b) conditions at which the plan stops to be executed (the plan is executed, impracticable or irrelevant on conditions); (c) actions which are carried out at a team level as a part of a common plan.

The assignment of roles and allocation of plans between the agents is carried out in two stages: at first the plan is distributed in terms of roles, and then the agent is put in correspondence to each role. One agent can execute a set of roles. Agents can exchange roles in dynamics of the plan execution. Requirements to each role are formulated as union of requirements to those parts of the plan which are put in correspondence to the role. There are also group and individual roles. Leaves correspond to individual roles. Agents' functionalities are generated automatically according to the roles.

For setting the agents' team operation in real-time a hierarchy of state machines is used. The state macbines realize a choice of the plan which will be executed and a fulfillment of the established sub-plans in a cycle "agents' actions - responses of environment".

At joint performance of the scenario agents' coordination is carried out by message exchange. As the agents' team fiunction in antagonistic environment agents can fail. Restoration of lost functionalities is carried out by means of redistribution of roles of the failed agent between other agents and cloning of new agents.

3. ONTOLOGY OF DDOS ATTACKS. STRUCTURE AND OPERATION OF AGENTS

The developed common ontology of DDoS attacks comprises a hierarchy of notions specifying activities of team of malefactors directed to implementation of attacks in different layers of detail. In this ontology, the hierarchy of nodes representing notions splits into two subsets according to the *macro- and micro-layers* of the domain specifications. All nodes of tbe ontology of DDoS attacks on the macro- and micro-levels of specification are divided into the *intermediate* and *terminal* (Kotenko and Man'kov 2003).

The notions of the ontology of an upper layer can be interconnected with the corresponding notions of the lower layer through one of three kinds of *relationships*: "*Part of*" that is decomposition relationship ("*Whole*"–"*Part*"); "*Kind of*" that is specialization relationship ("*Notion*"–"*Particular kind of notion*"); and "*Seq of*" that is relationship specifying sequence of operation ("*Whole operation*" – "*Sub-operation*").

High-layer notions corresponding to the intentions form the upper layers of the ontology. They are interconnected by the "*Part of*" relationship. Attack actions realizing malefactor's intentions (they presented at the lower layers as compared with tbe intentions) are interconnected with the intentions by "*Kind of*" or "*Seq of*" relationship.

The "terminal" notions of the macro-level are further elaborated on the *micro-level of attack specification*, and on this level they belong to the set of top-level notions detailed through the use of the three relationships introduced above.

In micro specifications of the computer network attacks ontology, besides the three relations described ("*Part of*", "*Kind of*", "*Seq of*"), the relationship "*Example of*" is also used. It serves to establish the "type of object–specific sample of object" relationship.

The developed ontology includes tbe detailed description of the DDoS domain in which the notions of the bottom layer ("*terminals*") are specified in terms of network packets, OS calls, and audit data.

Nodes specifying a set of software exploits for generation of DDoS attacks (Trinity V3, MSTREAM, SHAFT, TFN2K, Stacheldraht, Trin00) make up a top level of the ontology fragment. At lower levels different classes of DoS-attacks are detailed, for example "... flood" (sending a huge number of network packets wi... parameter). "Land" attacks (sending an IP-packet with ... fields of port and address of the sender and the recei... Source Address = Destination Address, Source Port N... = Destination Port Number), "Smurf" (sending broad... ICMP ECHO inquiries on behalf of a victim host. the ... hosts accepted such broadcasting packages answer ... victim host, that results in essential capacity reducti... communication channel or in full isolation of an a... network), etc.

DDoS-attack includes three stages: (1) preliminary, (2 ... and (3) final.

Main operations of the preliminary stage are inves... (reconnaissance) and installation of agents-"zombies". The content of the basic stage is realization of DDoS ... by joint actions of agents "master" and "daemons".

Common formal plan of attacks implemented by t... malefactors-agents has three-level structure:

(1) Upper level is a level of intention-based scena... malefactors' team specified in terms of seque... intentions and negotiation acts;

(2) Middle level is a level of intention-based scena... each malefactor specified in terms of ordered seque... sub-goals;

(3) Lower level is a level of malefactor's in... realization specified in terms of sequences of lo... actions (commands).

Algorithmic interpretation of the attack plan specifie... family of state machines. The basic elements of eac... machine are states, transition arcs, and explanatory te... each transition.

States of each state machine are divided into three ... first (initial), intermediate, and final (marker is End... initial and intermediate states are the following:

(1) non-terminal, those tbat initiate the work ... corresponding nested state machines;

(2) terminal, those that interact with the host model;

(3) abstract (auxiliary) states.

Example of one of realizations of the state machine ... represented in Figure 1. Main parameters of this rea... of the state machine are defined in Table 1.

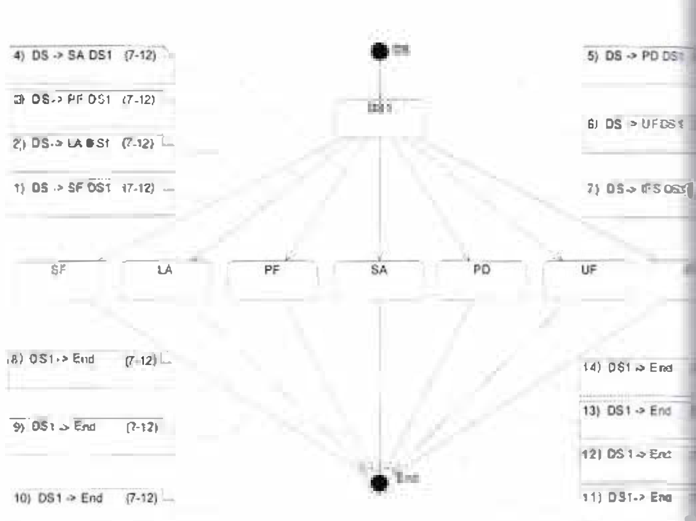


Figure 1: Diagram of State Machine DS (DoS attack)

[illegible] Parameters of State Machine DS

[illegible] name	DS
[illegible]	DS1, SF (SYN flood), LA (Land attack), PF (Ping flooding), SA (Smurf), PD (Ping of Death), UF (UDP flooding), IFS (Storm of inquiries to FTP-server), End
[illegible]	DS1
[illegible] states	-
[illegible] states	SF, LA, PF, SA, PD, UF, IFS
[illegible] states	DS1

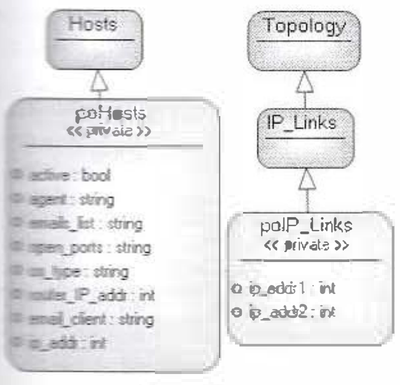
[illegible] of agents by two-level structure. The team [illegible] of "master" and "daemons". Each master [illegible] group of "demons". "Demons" execute [illegible] actions against victim hosts.

[illegible] development depends on the malefactor's "skill", [illegible] regarding network characteristics, which he/she [illegible] other malefactor's attributes. An attack is [illegible] as interactive process, in which the network [illegible] an attack action. Computer network plays the [illegible] environment for DDoS agents, and therefore its [illegible] be a part of the attack simulation tool.

[illegible] appears need in one more agent – agent-[illegible]". It simulates an attack environment – Internet. [illegible] is considered as a set of connected hosts. Every [illegible] characteristics, for example, ip-address and [illegible] hosts. All DDoS agent requests to outer world [illegible] by "simulator".

[illegible] is represented as follows: $a_N = <K, B, R, P, G,$ [illegible] – agent identifier; K – agent knowledge; B – [illegible] R – agent intentions; P – a set of parameters [illegible] the agent activity; $G = \{ L_R, f_R \}$ – a set of goals [illegible] L_R – hierarchy of possible goals and actions [illegible] influences), f_R – of choosing the goal or action [illegible] current sets K, B, R, P and C; C – [illegible] to other agents.

[illegible] (K) of the agent "simulator" is the [illegible] about hosts (active or not; ip-address; e-mails [illegible] open ports; OS type; ip-address of router for this [illegible] e-mail client, etc.) stored in the notion [illegible] the information about network topology (as a [illegible] represented in the notion "poIP_Links" [illegible]



[illegible] 2: Fragment of Agent "Simulator" Ontology

The agent-"simulator" beliefs (B) represent the information about agent deployed on the current host. It is stored in the notion "poHosts" (agent name).

The "simulator" parameters are the network topology and its hosts properties.

The agent set of goals and activities (L) consists of responses to other agents requests (they function according to protocols). The requests are: request for determining if the host active is (ActiveHostQuery); request for scanning the hosts ports (OpenPortQuery); request for host capturing (HostCapturing); DoS attack execution (DosExecution).

The commitments (C) to other agents are specified and fulfilled according to the protocols of interaction between the agents with defined roles (Figure 3).

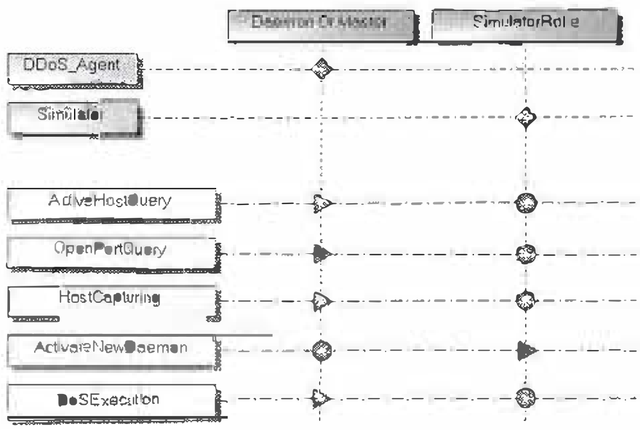


Figure 3: Agents, Their Roles and Protocols

The knowledge (K) of the agents "master" and "daemon" consists of information about compromised hosts and itself (one instance of the notion "poHost" and one "poAgentProps") (Figure 4).

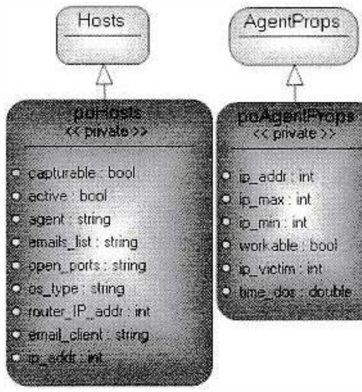


Figure 4: Fragment of Agent "Master" ("Daemon") Ontology

The agents "master" and "daemon" beliefs (B) are the information about environment (network topology, hosts parameters) and about other agents activity. They use the notions "poHosts" ("agent" attribute) and "poAgentProps" (host ip-address; ip-address range to scan; workable or not; victims ip-address; time to start attack). While acquiring new beliefs these agents build "the map of the world" step by step.

The agents parameters (*P*) are their activation parameters stored in the notion "AgentProps".

If the range of ip-addresses to scan is empty then the agent plays the "daemon" role. It executes the attack only. If the range is not null then the agent plays the "masters" role. It gathers the information about hosts from mentioned range, tries to capture them and also executes the attack.

The set of goals and activities (*G*) and their hierarchy (*L*) are represented by state machine representation (Figure 5).

Main goals and activities of "master" are as follows: Starting on accessible host ("DaemonActivation"); Gathering information about other hosts ("InformationGathering"), including Host activity determination ("GetActiveInfo") and Open ports determination ("GetOpenPortInfo"); Propagation by host capturing ("HostCapturing"), including Acquiring the hosts resources using the "shared resources" vulnerability ("Shared_Resources"); DoS attack execution ("AttackExecution"), for example using "Ping Of Death", "Syn overflow", "Smurf", etc. Main goals and activities of "daemon" are starting on accessible host and DoS attack execution.

The commitments (*C*) to other agents are specified and fulfilled according to the protocols of interaction between the agents with defined roles.

4. ATTACK SIMULATOR PROTOTYPE AND ITS EVALUATION

The software prototype of Attack Simulator has been implemented. Now it is used for validation of the accepted basic ideas, formal framework and implementation issues. The developed architecture of the attack simulator implementing the above described attack model was built as an agent of multi-agent system (MAS). The design and implementation of the attack simulator is being carried out on the basis of MAS DK – Multi-Agent System Development Kit (Gorodetski et al. 2002).

The MAS agents generated by MASDK have the same state-machine based architecture. Differences are reflected in the content of particular agents data and knowledge bases. Each agent interacts with other agents, environment which is perceived, and, possibly, modified by agents, and user communicating with agents through his interface.

The main objective of the experiments with the Attack Simulator prototype is to evaluate the tool's efficiency for different variants of attacks and attacked network configurations. These experiments were carried out for various parameters of the attack task specification and an attacked computer network configuration. The influence of the following input parameters on attacks efficacy was explored: a malefactor's intention, a degree of protection afforded by the network and personal firewall, a degree of security of attacked host, and the degree of malefactor's knowledge about a network. To investigate the Attack

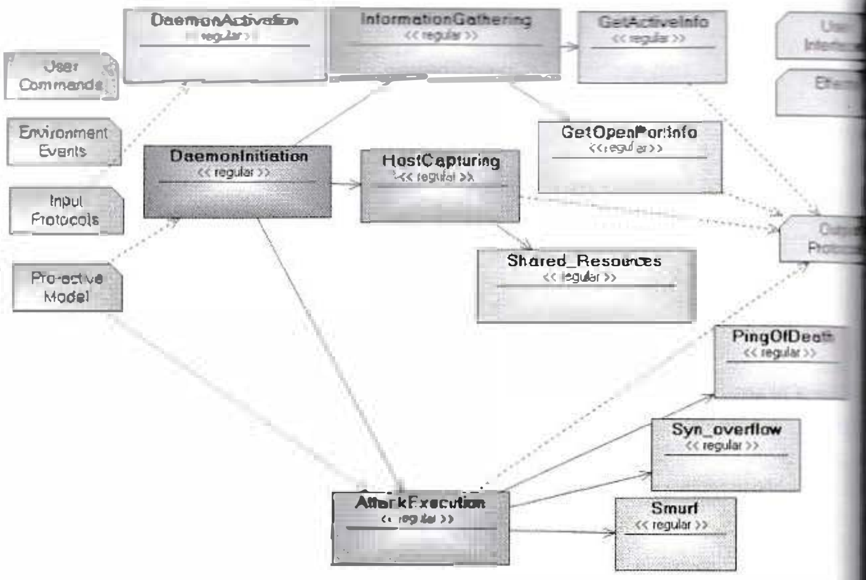


Figure 5: Fragment of "Master" ("Daemon") Goals and Activities

Simulator capabilities, the following parameters of realization outcome have been selected: number of te[rminal] level attack actions, percentage of the malefactor's in[tentions] that are successful, percentage of "effective" [...] responses on attack actions, percentage of attack [...] blockage by firewall, and percentage of "ineffective" [...] of attack actions.

Let us consider a small example of simulation of [...] attacks. The network fragment including 7 hosts def[...] the environment for DDoS is represented in Figure [...] Table 2.

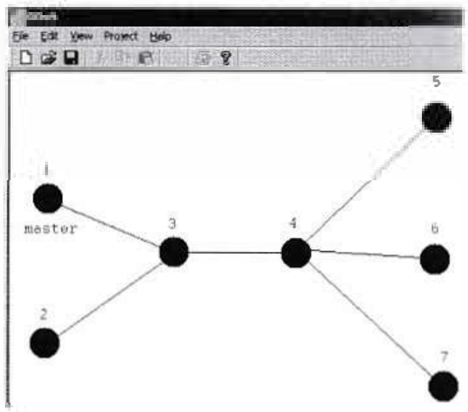


Figure 6: Graphic user interface for DDoS simula[tion]

Table 2: Initial conditions for simulation

№	IP address	Router IP-address	Active	Open ports	OS type	A[...]
1	1	3	yes	80	Win	ma[...]
2	2	3	yes	80,139	Win	-
3	3	-	yes	80	Win	-
4	4	-	yes	80,139	Win	-
5	5	4	yes	80	Win	-
6	6	4	yes	139	Win	-
7	7	4	yes	80	Win	-

[illegible] was deployed in the initial moment on the [illegible] parameters (*P*) were as follows: target of attack – [illegible] to compromise – 2-6; time to attack – 30 [illegible] after the start of simulation. Based on this data it [illegible] to create one instance of "simulator", one [illegible] of "master" and six instances of "daemons".
[illegible] logs its actions to the text file to trace the DDoS [illegible]. A part of DoS agent log is determined below:

```
[illegible] DDoS_Agent Master - Info gathering
[illegible] DDoS_Agent Master - SendMsg Ping (Active Query) ip=6
[illegible] DDoS_Agent Master - ReceiveMsg Re_Ping (Active
[illegible] ip=6, active=1
[illegible] DDoS_Agent Master - SendMsg PortScan (Open Ports
[illegible]
[illegible] DDoS_Agent Master - ReceiveMsg Re_IsPortOpen (Open
[illegible]) ip=6, open ports: 139
[illegible] DDoS_Agent Master - Capturing
[illegible] DDoS_Agent Master - Capturing: Shared Resources ip = 6
[illegible] DDoS_Agent Daemon4 - ReceiveMsg ActivateIt
[illegible] time_dos=1101135734000000
[illegible] DDoS_Agent Daemon2 - Attack Execution method= Ping
[illegible]
[illegible] DDoS_Agent Daemon3 - Attack Execution method= Smurf
[illegible] DDoS_Agent Daemon4 - Attack Execution method= Syn
[illegible]
```

[illegible] moment (18:01:44) the "simulator" (see data [illegible] 2) and "master" (see (P) above) were initialized. [illegible] began to gather information. It try to find if [illegible] hosts active and if they have the open ports. He [illegible] capture these hosts and to deploy "daemons" on [illegible]. They have waited until given time for attack [illegible]. As a result, "master" could capture the hosts with [illegible] 2, 4, 6 because they were active and had open [illegible]. So, only 4 agents (3 "daemons" and "master") [illegible] the attack on victim host (ip-address 7). The [illegible] DDoS method and attacked the host #7.
[illegible]-based exploration of the developed Attack [illegible] prototype has demonstrated its efficacy for [illegible] various attack scenarios against networks [illegible] structures and security policies implemented.

[illegible]SION

[illegible] we developed basic ideas of the modeling and [illegible] of DDoS attacks by teamwork approach. We [illegible] the structure of a team of agents, agent interaction-[illegible] mechanisms, and specifications of [illegible] of agent plans. The technology for creation of the [illegible] agents' team was suggested and described. We [illegible] the approach to be used for conducting [illegible] to both evaluate computer network security and [illegible] the efficiency and effectiveness of security policy [illegible] DDoS attacks. Software prototype of Attack [illegible] was developed. The attack simulator allows [illegible] a wide spectrum of real life DDoS attacks. Its [illegible] code is written in terms of Visual C++ and [illegible]. Experiments with the Attack Simulator have been [illegible], including the investigation of attack scenarios [illegible] networks with different security policies.
[illegible] development of the Attack Simulator tool will [illegible] enlargement of its capabilities in specification of [illegible] tasks, expansion of the DDoS attack classes [illegible] more sophisticated attack scenarios, etc.

11. ACKNOWLEDGEMENT

This research is being supported by grants 04-01-00167 of Russian Foundation of Basic Research and by the EC as part of the POSITIF project (contract IST-2002-002314).

REFERENCES

Cohen, P.R. and H.J. Levesque. 1991. "Teamwork". *Nous*, 25(4).

Fan, X. and J.Yen. 2004. "Modeling and Simulating Human Teamwork Behaviors Using Intelligent Agents". *Journal of Physics of Life Reviews*, Vol. 1, No. 3.

Gorodetski, V.; O. Karsayev; I. Kotenko; A. Khabalov. 2002. "Software Development Kit for Multi-agent Systems Design and Implementation". *Lecture Notes in Artificial Intelligence*, Vol. 2296.

Grosz, B. and S. Kraus. 1996. "Collaborative plans for complex group actions", *Artificial Intelligence*, Vol.86.

Jennings, N. 1995. "Controlling cooperative problem solving in industrial multi-agent systems using joint intentions". *Artificial Intelligence*, No.75.

Kotenko, I.; A. Alexeev; E. Man'kov. 2003. "Formal Framework for Modeling and Simulation of DDoS Attacks Based on Teamwork of Hackers-Agents". *IEEE/WIC International Conference on Intelligent Agent Technology*. Halifax, Canada, Proceedings. IEEE Computer Society.

Kotenko, I. and E. Man'kov. 2003. "Agent-Based Modeling and Simulation of Computer Network Attacks". *Fourth International Workshop "Agent-Based Simulation 4 (ABS 4)"*. Proceedings. Montpellier. France.

Mirkovic, J.; J. Martin; P. Reiher. 2002. "*A Taxonomy of DDoS Attacks and DDoS Defense Mechanisms*". Technical report #020018. University of California, Los Angeles.

Mirkovic, J.; S.Dietrich; D.Dittrich; P.Reiher. 2004. "*Internet Denial of Service: Attack and Defense Mechanisms*". Prentice Hall PTR.

Tambe, M. 1997. "Towards Flexible Teamwork". *Journal of Artificial Intelligence Research*, No.7.

Tambe, M. and D.V. Pynadath. 2001. "Towards Heterogeneous Agent Teams". *Lecture Notes in Artificial Intelligence*, Vol.2086.

BIOGRAPHY

IGOR KOTENKO graduated with honors from St.Petersburg Academy of Space Engineering and St.Petersburg Signal Academy. He obtained the Ph.D. degree in 1990 and the National degree of Doctor of Engineering Science in 1999. He is Professor of computer science and Leading researcher of St. Petersburg Institute for Informatics and Automation. His e-mail address is ivkote@spiiras.nw.ru and his Web-page can be found at http://space.iias.spb.su/ai/kotenko/ .

ALEXANDER ULANOV graduated from St. Petersburg State Politechnical University (2004), received his master's degree (2004) in the area "System analysis and control". He is now PhD student in the field of agent-based modeling and simulation for computer network attacks. His e-mail address is ulanov@iias.spb.su and his Web-page can be found at http://space.iias.spb.su/ai/ulanov/ .

Agent-based Artificial Consumer Market and Optimization of Defensive Strategies

Jürgen Wöckl[*†]

Abstract

This paper deals with a simulation approach to explore optimal defensive strategies concerning an entrance of a new firm in a gradually saturated consumer market. The question is how to react on a new entrance from the view of the incumbents and also how to succeeds entering the market from the view of the new entrant.

A agent-based artificial consumer market (AB-ACM) has been designed to provide an experimental environment for the optimization task. The AB-ACM has been formulated very generally to make sure all interested marketing and management related strategies can be considered. Optimization methods are applied to the artificial market to derive optimal strategies obtaining maximal profits to keep competitive and to find optimal reactions for both - the entrant and the incumbents in the market. In the concluding simulation study there is one questions of interest concerning the adaptation of the strategy to obtain maximal profits - the effect of varying heterogeneity of the considered consumer aspiration points. The resulting optimal price-budget combinations remain stable up to a certain degree of heterogeneity. Then a threshold is reached from which further disaggregating the market will lead to boundary solutions.

1 Introduction to the AB-ACM

The artificial market is made up of a constant number of consumers which are represented by artificial agents creating a agent-based framework. Each consumer has an individual aspiration point of attributes which the preferred product should possess. The choice process of the artificial consumer depends on the knowledge of the products offered at the market and the attitude the consumer gained by comparing his aspiration point to the perceived attributes of the product (see [Buchta and Mazanec(2001)]). In the environment it is assumed that each firm just provides one product and therefore the profit of the firm equals the price times sales of its product reduced by the budget spent for advertising. Additionally the variables price and budget are relevant for the attitude of a consumer regarding to the product. The attitude depends to the advertising budget of a firm weighted by the price of the product. At the initial time ($t_0 = 0$) no consumer agent knows anything about the products and the firms on the market, and their choices cannot be made rationally - in the context of choosing a product best fullfilling their aspirations. Primary through the advertising of the firms the consumers get information about the products and their attributes and so they are able to choose the best fitting product. The success of each firm/product depends on the price, the attributes and the invested advertising budget. These variables has to be optimized by all firms in the market, especially at the time of entry of a new firm attacking a segment allready covered by an incumbent.

2 Functionality of the AB-ACM

The following equation (1) shows the evolvement of the attitudes att_{ijk} of a consumer i regarding to the attribute k of product j. This differential equation consists of two parts where the first describes the growth of the attitudes of the advertised attributes starting at 0 up to 1 dependent on the actual relative advertising budget. The second part describes the decay due to the forgetting of the product attributes by the consumers. The appropriate function $b(\cdot)$ is defined in equation (2) and (3).

In the following the indices i denote the aspiration groups, j the brands, k the product attributes and t the time.

The differential equation responsible for the temporal modification of the attitudes p of those attributes which are advertised is the following:

$$\frac{datt_{ijk}(t)}{dt} = \frac{1}{price_j^*} \cdots \cdot \left[aif(budget_j)(1 - att_{ijk}(t)) - b(t, budget_j)att_{ijk}(t)\right] \quad (1)$$

where $aif(budget_j)$ indicates the advertising impact function depending to the advertising budget of product j.

*Adaptive Methodology and Corporate Learning; WU-Wien; juergen.woeckl@wu-wien.ac.at

†ARGESIM; TU Wien; jwoeckl@osiris.tuwien.ac.at

$budget_j$) characterize the forgetting rate of the consumer and $price^*_j$ refer the relative price of product j:

$$aif(budget_j) = e^{\alpha - \frac{\beta}{budget_j}}$$

$$price^*_j = \frac{price_j}{\frac{1}{J}\sum_{j=1}^{J}(price_j)}$$

In this equation the oblivion rate is described by:

- for non-advertised attribute:

$$b(t, budget_j) = b_0 \tag{2}$$

- for advertised attribute:

$$b(t, budget_j) = \frac{1}{1 + \mathcal{F}(t, budget_j)} \tag{3}$$

with

$$\begin{aligned} \mathcal{F}(\cdots) &= budget_j(t) \cdots \\ & \int_0^t \frac{budget_j(\tau)}{\sum_j (budget_j(\tau))} \cdots \\ & \cdot e^{-b_0 (t-\tau)} d\tau \end{aligned} \tag{4}$$

Equation (3) describes the advertising effect of budget spent in the past. The amount of all advertising budgets affects the present consumers' attitude. The factor $aif(budget_j)$ in the ODE (equ. 1) regulates the latency of advertising effects - in other words the oblivion of the consumer. So alternating advertising strategies are successful because of the nonlinear continuation of the effect of former advertising budgets. If the advertising is stopped after some time the effect vanishes and the oblivion rate of the consumers regarding the specific product increases and the attitude decreases continuously. The utility of the consumer i with respect to each product j can be measured using the proportional distance between the appropriate aspiration point and the attitude corresponding to brand j and is calculated as

$$uti_{ij} = \frac{max(distance_{ij})}{distance_{ij}} \tag{5}$$

3 Design of the Agent-based Environment

The agent-based environment has been implemented using Python and Matlab and the object oriented structure of the agents is shown in fig. (1). The number of agents and firms are constant and for the experiments we used 300 consumer agents and 4 firm agents. The consumers are divided into three equal segments, each containing 100 consumers. The three segments are well separated and all the consumer aspiration points of a segment are normally distributed around a segment specific common ideal point. At initial time $t_0 = 0$ three of the firm agents are in the market and after some time the steady state of each firm serves one segment has arose. This follows because in this state each firm earns sufficient money from the 100 consumers they are serving exclusively and beetween the firms there is no competition necessary. After the market is in this stable but still partly saturated state, the fourth firm enters the market and attacks the segment of one of the incumbents. The market is fully saturated if all consumers have perfect information about all firms/products. The saturation of the market is one of the design factors of this study. In the next section the design and the hypothesis of the study are presented in more detail.

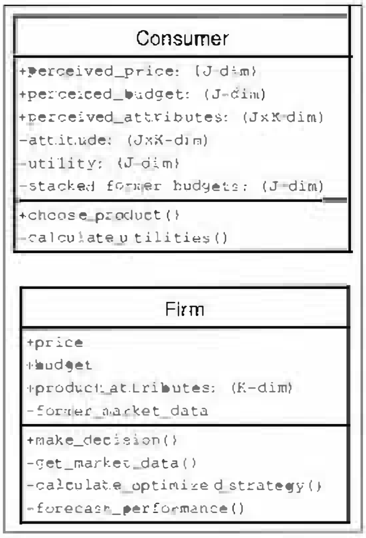


Figure 1: Class Diagrams of the agents 'Consumer' and 'Firm'

4 Experiments and Results

The market which in this study is assumed as an exemplary application for the optimization task consists of three well separated consumer segments each of them treated by a single firm. To optimize their individual profits the brands are able to set the price of the product and the advertising budget to evoke higher consumer preferences and thus higher market shares. After a period of increasing market saturation a new firm enters the market in a specific segment and thus attacks the position of

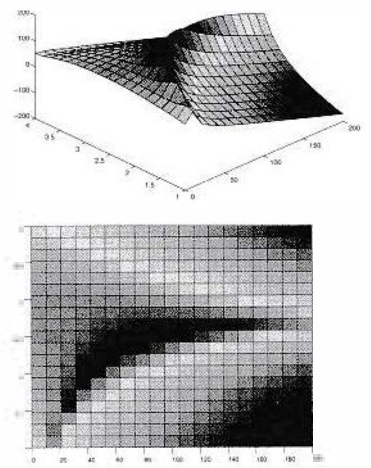


Figure 2: Surface plots of the profits of the incumbent for several price-budget combinations under a fixed entrance strategy in a homogeneous market ($\sigma^2 = 0$)

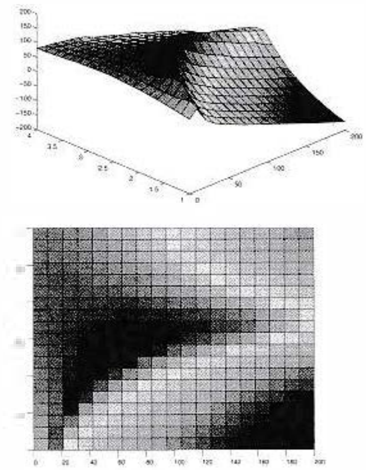


Figure 3: Surface plots of the profits of the incumbent for several price-budget combinations under a fixed entrance strategy in a heterogeneous market ($\sigma^2 = 0.04$)

one incumbent. The emphasis in this study is to derive the optimal reaction of the incumbent as well as the optimal entrance strategy concerning prices and advertising expenditures. The diversification of those optimal strategies varying a certain experimental factor is presented. Here the factor of interest is the heterogeneity of the considered consumer aspiration points. It is expected that the results of the experiment will become more fuzzy with increasing heterogeneity. The firms' target is to optimize their profit, especially after the new entrant participates the market the achievement of an equilibrium is very ambitious. The following figure shows a exemplary plot of the target function for the optimization routines. More results and their interpretation are discussed in the article.

4.1 Optimization of the incumbent varying the "Heterogeneity of the consumers' aspirations"

The experimental factor considered concerns the degree of heterogeneity in each consumer segment. There can either be a single ideal point per segment (homogeneous case) or a more or less diverging pall of individual consumer ideal points (heterogeneous case).

As a first assumption the market segments are homogeneous. That means though each consumer segment consists of 100 consumers, they are assumed to have the same aspiration point. This common aspiration level can be interpreted as the segment ideal point ($\sigma^2 = 0$). Further the incumbents' reaction in case of a fixed entrance strategy is investigated under the assumption that the individual ideal points of the consumers are normally distributed with the segment ideal point as mean and a relatively small variance ($\sigma^2 = 0.04$ and $\sigma^2 = 0.08$) in each segment respectively (fig. 3 and 4). The surface plots are served for all three cases to present the resulting profits for the incumbent under a fixed entrance strategy.

As it can be seen in Figure 2 there is no unique optimum. There exist different price-budget combinations resulting in the same optimal profit for the incumbent. But at least a general tendency concerning price and budget reactions can be read off. More precise the incumbent should reduce its price (down to a value between 2 and 2.5 units) as well as its advertising budget (down to a value in the interval [50, 100]) when it is facing a new brand which is entering the market and directly jeopardizing its monopoly in this special market segment.

For the case of distributed ideal points (fig. 3 and 4) the tendency of the prices to stay constant and the budgets to decrease still holds. Prices should be selected from the interval $[2, 2.5]$ and budgets out of $[40, 60]$. Another interesting result the tendency towards a boundary solution. Boundary values like stop spending

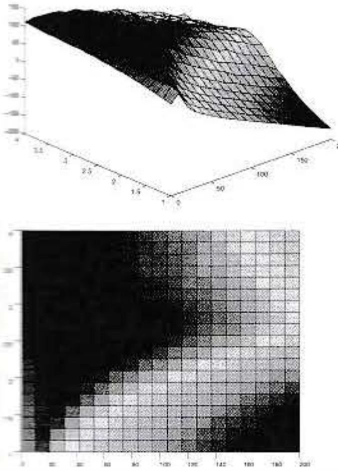


Figure 4: Surface plots of the profits of the incumbent for several price-budget combinations under a fixed entrance strategy in a heterogeneous market ($\sigma^2 = 0.08$)

advertising expenditures at all and set the highest price possible become more and more attractive. For further application studies and a more detail discussion of the strategic suggestions see [Schuster and Wöckl(2005)] and [Wöckl and Schuster(2004)]. Further literature about agent-based implementations and the general approach to the used model can be found in [Buchta and Mazanec(2001)] and [Schwaiger and Stahmer(2003)].

References

[Ansari et al.(1994)Ansari, Economides, and Ghosh] Asim Ansari, Nicholas Economides, and Avijit Ghosh. Competitive Positioning in Markets with Nonuniform Preferences. *Marketing Science*, 13(3):248–273, Summer 1994.

[Basuroy and Nguyen(1998)] Suman Basuroy and Dung Nguyen. Multinomial Logit Market Share Models: Equilibrium Characteristics and Strategic Implications. *Management Science*, 44(10):1396–1408, October 1998.

[Buchta and Mazanec(2001)] Christian Buchta and Josef Mazanec. SIMSEG/ACM A Simulation Environment for Artificial Consumer Markets. Working Paper 79. SFB - Adaptive Modelling: www.wu-wien.ac.at/am, March 2001.

[Carpenter and Nakamoto(1990)] Gregory S. Carpenter and Kent Nakamoto. Competitive Strategies for Late Entry into a Market with a Dominant Brand. *Management Science*, 36(10):1268–1278, October 1990.

[Gatignon et al.(1989)Gatignon, Anderson, and Helsen] Hubert Gatignon, Erin Anderson, and Kristiaan Helsen. Competitve Reactions to Market Entry: Explaining Interfirm Differences. *Journal of Marketing Research*, 26:44–55, February 1989.

[Gruca et al.(2001)Gruca, Sudharshan, and Kumar] Thomas S. Gruca, D. Sudharshan, and K. Ravi Kumar. Marketing Mix Response to Entry in Segmented Markets. *International Journal of Research in Marketing*, 18:53-66, 2001.

[Hauser and S.(1983)] John R. Hauser and Shugan Steven S. Defensive Marketing Strategies. *Marketing Science*, 2(4):319 – 360, 1983. Fall.

[Hruschka(1996)] Harald Hruschka. *Marketing-Entscheidungen*. Vahlen, 1996.

[Kumar and Sudharshan(1988)] K. Ravi Kumar and D. Sudharshan. Defensive Marketing Strategies: an Equlibrium Analysis based on Decoupled Response Function Models. *Management Science*, 34(7):805–815, July 1988.

[Lilien et al.(1992)Lilien, Kotler, and Moorthy] Gary Lilien, Philip Kotler, and Sridhar Moorthy. *Marketing Models*. Prentice Hall International, 1992.

[Lilien and Rangaswamy(2003)] Gary L. Lilien and Arvind Rangaswamy. *Marketing Engineering*. Addison-Wesley, 2003.

[Schuster and Wöckl(2005)] Ulrike Schuster and Jürgen Wöckl. Optimal Defensive Strategies under Varying Consumer Distributional Patterns and Market Maturity. *Journal of Economics and Management*, 1(2): 187 – 206, July 2005.

[Schwaiger and Stahmer(2003)] Arndt Schwaiger and Björn Stahmer. Simmarket: Multiagent-based customer simulation and decision support for category management. In *MATES 2003*, LNAI 2831, pages 74–84. Springer, 2003.

[Wöckl and Schuster(2004)] Jürgen Wöckl and Ulrike Schuster. Derivation of stationary optimal defensive strategies using a continuous market model. In *AMS Annual Conference*, volume XXVII, pages 305–311, 2004.

INTEGRATIVE NEGOTIATION AMONG AUTONOMOUS AGENTS

Fernando Lopes, A. Q. Novais
INETI, DMS
Est. do Paco do Lumiar
1649-038 Lisboa, Portugal
{fernando.lopes, augusto.novais@ineti.pt}

Nuno Mamede
IST, L²F
R. Alves Redol 9
1000-029 Lisboa, Portugal
Nuno.Mamede@inesc-id.pt

Helder Coelho
UL, FC, DI
Campo Grande
1749-016 Lisboa, Portugal
hcoelho@di.fc.ul.pt

KEYWORDS

Autonomous Agents, Conflict of Interests, Negotiation.

ABSTRACT

Autonomous agents generate plans towards the achievement of their goals and, over time, situations arise in which their plans conflict with the plans of other agents. Negotiation is the predominant process for resolving conflicts. This paper presents the key features of a negotiation model for autonomous agents that handles multi-party, multi-issue and repeated rounds. The model acknowledges the role of conflict as a driving force of negotiation, formalizes a set of human negotiation procedures, allows the dynamic addition and removal of issues, and accounts for a tight integration of the individual capability of planning and the social capability of negotiation. This paper also describes an experiment conducted to evaluate a version of the model that handles two-party, multi-issue negotiation. The results confirmed a number of conclusions about human negotiation.

INTRODUCTION

Autonomous agents generate plans towards the achievement of their goals. The agents operate in complex environments and situations often arise in which their plans conflict with the plans of other agents. The predominant process for resolving conflicts is negotiation. Recent growing interest in electronic commerce and supply chain management has given increased importance to negotiation.

This paper presents the key features of a generic model of negotiation that handles multi-party, multi-issue and repeated rounds. The main components of the model are: (i) a prenegotiation model, (ii) a multilateral negotiation protocol, (iii) an individual model of the negotiation process, (iv) a set of negotiation strategies, and (v) a set of negotiation tactics. The model accounts for a tight integration of individual and social behavior. Also, the model acknowledges the role of conflict as a driving force of negotiation, formalizes a set of human negotiation procedures, and allows the dynamic addition and removal of issues.

The model is currently being evaluated. This paper describes an experiment conducted to: (i) assess the feasibility of building autonomous negotiating agents equipped with a version of the model that handles two-party, multi-issue negotiation (integrative negotiation), (ii) investigate the behavior of integrative strategies and their associated tactics, and (iii) evaluate the effect of these strategies and tactics both on the process and on the outcome of negotiation.

This paper builds on our previous work in the area of negotiation (Lopes et al. 2002; Lopes et al. 2004). The remainder of the paper is structured as follows. A generic model of individual behavior for autonomous agents is presented first, followed by the key features of a generic model of negotiation. Next, the experimental work is reported. Finally, related work and concluding remarks are presented in the last two sections.

AUTONOMOUS AGENTS

This section presents a model of individual behavior that captures some important features of a wide range of autonomous agents. Let *Agents* be a set of agents. A brief description of the key features of every agent $ag_i \in Agents$ follows (see also Lopes et al. 2002).

The agent ag_i has a set $B_i=\{b_{i1},\ldots\}$ of beliefs, a set $G_i=\{g_{i1},\ldots\}$ of goals, and a library $PL_i=\{pt_{i1},\ldots\}$ of plan templates. *Beliefs* represent information about the world and the agent himself, *goals* represent world states to be achieved, and *plan templates* are procedures for achieving goals. Every plan template $pt_{ij} \in PL_i$ is a 6-tuple that includes a header, a type, a list of conditions, a body, a list a constraints, and a list of statements. The library PL_i has *composite* plan templates specifying the decomposition of goals into more detailed subgoals, and *primitive* plan templates specifying actions directly executable by ag_i.

The agent ag_i is able to generate complex plans from the simpler plan templates stored in the library. A *plan* p_{ik} for achieving a goal g_{ik} is a 3-tuple that includes a list $PT_{ik} \subseteq PL_i$ of plan templates, a binary relation that establishes a hierarchy on PT_{ik}, and another binary relation that establishes a temporal order on PT_{ik}. The plan p_{ik} is represented as a hierarchical and temporally constrained And-tree denoted by $Pstruct_{ik}$.

At any instant, the agent ag_i has a number of plans for execution. These plans are the plans adopted by ag_i and are stored in the *intention structure* $IS_i=[p_{i1},\ldots]$. For each plan $p_{ij} \in IS_i$, the header of every plan template pt_{ijm} in p_{ij}, is referred as *intention* int_{ijm}. The agent ag_i often has information about the other agents in *Agents*. This information is stored in the *social description* $SD_i=\{SD_i(ag_1),\ldots\}$.

The agent ag_i checks regularly its adopted plans in order to detect any potential conflict of interests. To this end ag_i has a library of *conflict detection axioms* $CL_i=\{ax_{i1},\ldots\}$. The axioms state which intentions cannot be satisfied together.

THE NEGOTIATION MODEL

Let $Ag=\{ag_1,\ldots,ag_n\}$, $Ag \subseteq Agents$, be a set of autonomous agents. Let $P_{Ag}=\{p_{1k},\ldots,p_{nk}\}$ be a set of plans of the agents in Ag including intentions $I_{Ag}=\{int_{1kl},\ldots,int_{nkl}\}$, respectively for agents $ag_1,\ldots,ag_n$. Let the intentions in I_{Ag} represent commitments to achieve exclusive world states. In this situation, there is a conflict $Conf_{Ag}$ among the agents in Ag. This section presents the key features of a negotiation model (see our earlier work for an in-depth discussion (Lopes et al. 2002; Lopes et al. 2004)).

Preparing and Planning for Negotiation

The prenegotiation model defines the main tasks that each agent $ag_i \in Ag$ must attend to in order to prepare for negotiation. A description of these tasks follows.

Generation of the Negotiation Problem Structure. A negotiation problem NP_{ik} from the perspective of ag_i is a tuple that includes a goal g_{ik}, a plan p_{ik} for achieving an intention int_{ikl} of p_{ik}, the set $A_i=Ag-\{ag_i\}$ of agents negotiating with ag_i, and the set of intentions $I_{A_i}=I_{Ag}-\{int_{ikl}\}$. The problem NP_{ik} has a structure $NPstruct_{ik}$ consisting of a hierarchical And-Or tree. The nodes of the And-Or tree are plan templates. The header of the root node describes the *negotiation goal* g_{ik}. The structure $NPstruct_{ik}$ defines all the solutions of NP_{ik} currently known by ag_i. A *solution* is a plan that can achieve the negotiation goal g_{ik}.

Issue Identification and Prioritization. The negotiation issues of ag_i are obtained from the leaves of $NPstruct_{ik}$. Let $L_{ik}=[pt_{ika},\ldots]$ be the collection of plan templates constituting the leaves of $NPstruct_{ik}$. The header of every plan template $pt_{ikj} \in L_{ik}$ is called a *fact* and denoted by f_{ikj}. Formally, a *fact* f_{ikj} is as 3-tuple that includes a negotiation *issue* is_{ikj} and a value $v[is_{ikj}]$ of is_{ikj}. Let $F_{ik}=\{f_{ika},\ldots,f_{ikz}\}$ be the set of facts of $NPstruct_{ik}$. The negotiating *agenda* of ag_i is the set of issues $I_{ik}=\{is_{ika},\ldots,is_{ikz}\}$ associated with the facts in F_{ik}. The interval of legal values for each issue $is_{ikj} \in I_{ik}$ is represented by $D_{ikj}=[min_{ikj},max_{ikj}]$. The *priority* of is_{ikj} is a number that represents its order of preference. The *weight* of is_{ikj} is a number that represents its relative importance. The sets of priorities and normalized weights of the issues in I_{ik} are represented by $PR_{ik}=\{pr_{ika},\ldots,pr_{ikz}\}$ and $W_{ik}=\{w_{ika},\ldots,w_{ikz}\}$, respectively.

Limits and Aspirations Formulation. The *limit* or *reservation value* is the value beyond which a bargainer is unwilling to concede. The *aspiration* is the value sought at any particular time. The limit for each issue is_{ikj} in I_{ik} is represented by lim_{ikj} and the initial aspiration by asp^{t1}_{ikj}.

Negotiation Constraints Definition. Negotiation constraints bound the possible values for the issues in I_{ik}. *Hard constraints* are linear constraints that specify threshold values for the issues. *Soft constraints* are linear constraints that specify minimum acceptable values for the issues.

Negotiation Strategy Selection. The agent ag_i has a library $SL_i=\{str_{i1},\ldots\}$ of negotiation strategies and a library $TL_i=\{tact_{i1},\ldots\}$ of negotiation tactics. The selection of a strategy is an important task and must be carefully planned (Pruitt and Kim 2004). In this paper, we just assume that ag_i selects a strategy $str_{ik} \in SL_i$ that it considers appropriate according to its experience.

The Multilateral Negotiation Protocol

The protocol specifies the set of possible tasks that the agents in Ag can perform during the negotiation process. A *global* description of this process follows.

The negotiation process starts with an agent, say ag_i, communicating a negotiation proposal $prop^{t1}_{ikm}$ to all the agents in A_i. A *negotiation proposal* is a set of facts (see next subsection). Each agent $ag_j \in A_i$ receives $prop^{t1}_{ikm}$ and may decide either: (i) to accept $prop^{t1}_{ikm}$, (ii) to reject $prop^{t1}_{ikm}$ without making a critique, or (iii) to reject $prop^{t1}_{ikm}$ and making a critique. A *critique* is a comment on which parts of a proposal are acceptable and unacceptable or a statement about relevant aspects of the negotiation process.

The process continues with ag_i receiving the responses of all the agents in A_i. Next, ag_i checks whether a negotiation agreement was reached. If the proposal $prop^{t1}_{ikm}$ was accepted by all the agents in A_i, the negotiation process ends successfully. In this case, ag_i just informs the agents in A_i that an agreement was reached. Otherwise, ag_i can act either: (i) by communicating a new proposal $prop^{t3}_{ikn}$, or (ii) by acknowledging the receipt of all the responses.

The process proceeds with the agents in A_i receiving the response of ag_i. If ag_i decides to communicate a new proposal $prop^{t3}_{ikn}$, each agent ag_j in A_i may again perform the tasks just specified. If ag_i decides to acknowledge the receipt of all the responses, the process proceeds to a new round in which another agent $ag_k \in Ag$ communicates a proposal to all the agents in $A_k=Ag-\{ag_k\}$. This is repeated for other agents in Ag.

The Negotiation Process: Individual Perspective

The individual model of the negotiation process defines the tasks that each agent $ag_i \in Ag$ can perform during the negotiation process. A brief description of these tasks follows (for simplicity, we omit the time).

Negotiation Proposal Generation. This process generates the set of initial proposals NPS_{ik} satisfying the requirements imposed by $NPstruct_{ik}$. The generation of NPS_{ik} is performed through an iterative procedure involving: (i) problem interpretation, (ii) proposal preparation, and (iii) proposal addition. Problem interpretation consists of searching $NPstruct_{ik}$ for any solution sol_{ikm} of NP_{ik} and selecting the primitive plan templates $PPT_{ikm}=\{pt_{ika},\ldots,pt_{ikp}\}$ of sol_{ikm}. Proposal preparation consists of determining a *negotiation proposal* $prop_{ikm}=\{f_{ika},\ldots,f_{ikp}\}$, i.e., a set of facts corresponding to the headers of the plans in PPT_{ikm}. Proposal addition consists of adding $prop_{ikm}$ to NPS_{ik}.

Feasible and Acceptable Proposal Preparation. This process generates the set of feasible proposals FPS_{ik}, $FPS_{ik} \subseteq NPS_{ik}$, and the set of acceptable proposals APS_{ik}, $APS_{ik} \subseteq FPS_{ik}$. Let $Iprop_{ikm} = \{is_{ika}, \ldots, is_{ikp}\}$ be the issues associated with the facts in $prop_{ikm} = \{f_{ika}, \ldots, f_{ikp}\}$. Also, let $HCprop_{ikm} = \{hc_{ika}, \ldots, hc_{ikp}\}$ and $SCprop_{ikm} = \{sc_{ika}, \ldots, sc_{ikp}\}$ be the sets of hard and soft constraints for the issues in $Iprop_{ikm}$, respectively. A negotiation proposal $prop_{ikm}$ is *feasible* if the issues in $Iprop_{ikm}$ satisfy the set $HCprop_{ikm}$ of hard constraints. A feasible proposal $prop_{ikm}$ is *acceptable* if the issues in $Iprop_{ikm}$ satisfy the set $SCprop_{ikm}$ of soft constraints.

Feasible Proposal Evaluation. This process computes a score for each proposal in FPS_{ik}. In this paper, we consider that the score of $prop_{ikm}$ is given by an *additive scoring function* (Raiffa 1982).

Feasible Proposal Selection. This process selects a feasible proposal $prop_{ikm} \in FPS_{ik}$. The strategy str_{ik} of ag_i defines a tactic $tact_{ik} \in TL_i$ to use. The tactic $tact_{ik}$ specifies a particular proposal $prop_{ikm}$ (see the next two subsections).

Feasible Proposal Modification. This process computes a new proposal from a rejected proposal $prop_{ikm}$. The strategy str_{ik} defines one or more tactics to use. The tactics modify $prop_{ikm}$ to make it more acceptable. The modification can be done either: (i) by making a concession, or (ii) without making a concession (see again the next two subsections).

Negotiation Strategies

Negotiation strategies are functions that define the tactics to be used at the beginning and during the course of negotiation. This subsection describes two classes of strategies, called concession and problem solving (or integrative) strategies.

Concession strategies. These strategies are functions that model well-known concession patterns (Lewicki et al 2003; Carnevale and Pruitt 1992). In this paper, we consider the following three sub-classes of strategies:

1. *starting high and conceding slowly* – model an optimistic opening attitude and successive small concessions;
2. *starting reasonable and conceding moderately* – model a realistic opening attitude and successive moderate concessions;
3. *starting low and conceding rapidly* – model a pessimistic opening attitude and successive large concessions.

The "starting high and conceding slowly" strategies are formalized by similar functions. For instance, a strategy $shslw\_01$ is formalized by the following function:

$shslw\_01(TL_i) = (class, tact_{ik})$ |

if: $state$ = "*initial*" then:

$class$ = "*opening_negotiation*" $\wedge$

$tact_{ik}$ = "*starting_optimistic*"

else:

$class$ = "*constant_concession_factor*" $\wedge$

$tact_{ik}$ = "*tough*"

where $state$ is the current state of the negotiation, $class$ denotes the class of the tactic $tact_{ik}$ specified by the strategy, and *starting_optimistic* and *tough* are tactics (see next subsection). The strategies in the other two-subclasses are formalized by similar functions.

Problem solving or integrative strategies. These strategies are functions that model negotiation procedures leading to integrative agreements. *Integrative or win-win agreements* are agreements that provide high joint benefit (Pruitt and Kim, 2004; Lewicki et al 2003). In this paper, we consider the following two sub-classes of strategies:

1. *low priority concession making* – model small concessions on issues of high priority and large concessions on issues of low priority;
2. *modified logrolling* – model large concessions both on issues of low priority for ag_i and on issues of high priority for the other agents;

The strategies in these sub-classes partition a set of issues, say $Iprop_{ikm}$, into: (i) subset $Iprop^{+}_{ikm}$, corresponding to higher priority issues, (ii) subset $Iprop^{-}_{ikm}$, corresponding to lower priority issues, and (iii) subset $Iprop^{\pm}_{ikm}$, corresponding to remaining issues.

The "low priority concession making" strategies are similar. For instance, a strategy $srmlp\_01$ that specifies a realistic opening attitude, small concessions on issues of high priority, large concessions on issues of low priority, and moderate concessions on the remaining issues, is formalized by the following function:

$srmlp\_01(Iprop_{ikm}, PR_{ik}, TL_i) = (class, tact_{ik}, Iprop^{+}_{ikm}, tact_{ik+1}, Iprop^{\pm}_{ikm}, tact_{ik+2}, Iprop^{-}_{ikm})$ |

if: $state$ = "*initial*" then:

$class$ = "*opening_negotiation*" $\wedge$

$tact_{ik}$ = "*starting_realistic*" $\wedge$ $tact_{ik+1} = tact_{ik+2}$ = "*nil*"

else:

$class$ = "*constant_concession_factor*" $\wedge$

$Iprop_{ikm} = Iprop^{+}_{ikm} + Iprop^{\pm}_{ikm} + Iprop^{-}_{ikm}$ $\wedge$

$\forall is_{ikj} \in Iprop^{+}_{ikm}$, $tact_{ik}$ = "*tough*" $\wedge$

$\forall is_{ikj} \in Iprop^{\pm}_{ikm}$, $tact_{ik+1}$ = "*moderate*" $\wedge$

$\forall is_{ikj} \in Iprop^{-}_{ikm}$, $tact_{ik+2}$ = "*soft*"

where $tact_{ik}$, $tact_{ik+1}$ and $tact_{ik+2}$ denote the tactics specified by the strategy, and *starting_realistic*, *moderate* and *soft* are tactics (see subsection "Negotiation Tactics").

The "modified_logrolling" strategies are also similar. Lopes et al. (2004) present a formal description of a strategy $srsml\_01$ that specifies an optimistic opening attitude, small concessions on issues of high priority, large concessions on issues of low priority, large concession on issues of moderate priority for ag_i (and high priority for the other agents), and small concessions on the remaining issues of moderate priority.

Negotiation Tactics

Negotiation tactics are functions that define the moves to be made at each point of the negotiation process.

Opening negotiation tactics. These tactics are functions that specify a proposal to submit at the beginning of negotiation. Let $FPS_{ik}=\{prop_{ik1},\ldots,prop_{ikw}\}$ and $APS_{ik}=\{prop_{ik1},\ldots,prop_{ikh}\}$ be the ordered sets of feasible and acceptable proposals of ag_i, respectively. Let $NAPS_{ik}=FPS_{ik}-APS_{ik}$. In this paper, we consider three tactics (for simplicity, we omit the time):

1. *starting optimistic* – specifies the proposal $prop_{ik1}$ with the highest score $Vprop_{ik1}$;
2. *starting realistic* – specifies either: (i) proposal $prop_{ikh}\in APS_{ik}$ with the lowest score, or (ii) proposal $prop_{ikh+1}\in NAPS_{ik}$ with the highest score;
3. *starting pessimistic* – specifies the proposal $prop_{ikw}$ with the lowest score $Vprop_{ikw}$.

Concession tactics. These tactics are functions that compute new values for each issue during the negotiation process. Let $is_{ikl}\in I_{ik}$ be an issue at stake in negotiation. In this paper, we consider a *constant concession factor* sub-class of tactics. In this sub-class, we consider five tactics:

1. *stalemate* – models a *null* concession on is_{ikl};
2. *tough* – models a *small* concession on is_{ikl};
3. *moderate* – models a *moderate* concession on is_{ikl};
4. *soft* – models a *large* concession on is_{ikl};
5. *compromise* – models a *total* concession on is_{ikl}.

EXPERIMENTAL ANALYSIS

This section describes an experiment aiming at: (i) assessing the feasibility of building autonomous agents equipped with a version of the negotiation model that handles two-party, multi-issue negotiation, and (ii) evaluating the integrative strategies and their associated tactics by confirming a number of basic conclusions about human negotiation.

Empirical Research on Human Negotiation

Negotiation is a rich research area. Most studies are laboratory experiments on bilateral negotiation about two or more issues of different priority (integrative negotiation). Experimental evidence supports the following two conclusions (Lewicki et al 2003; Carnevale and Pruitt 1992):

1. systematic trial and error, in which one or both parties frequently make new proposals, or concede systematically (*i.e.*, explore various options at each profit level before proceeding to a lower level), or make large concessions on low priority issues, promotes the development of integrative agreements;
2. information exchange, in which one or both parties provide information about their priorities or the interests underlying their positions, promotes the development of integrative agreements.

The Experimental System

The experimental system consists of two autonomous agents and an environment. Let $Ag=\{ag_s, ag_b\}$ be the set of agents. The agent ag_s plays the role of a seller and the agent ag_b the role of a buyer. The agents negotiate the price, down payment, financing terms and delivery date of a commodity denoted by $prod_X$. A description of the agents and the environment follows.

Autonomous negotiating agents. Every agent $ag_i\in Ag$ is equipped with the model of individual behavior described in section "Autonomous Agents". We consider the following (for simplicity, we drop the subscripts k and j):

- the set G_i of every agent $ag_i\in Ag$ contains the goal g_i of selling (or buying) $prod_X$;
- the library PL_i contains thirteen plan templates;
- the intention structure IS_i contains the plan p_i for achieving the goal g_i;
- the library CL_i contains one axiom.

Every agent ag_i is also equipped with a simplified version of the negotiation model described in the section "The Negotiation Model". The process of preparing and planning for negotiation involves the tasks just specified, except the task "negotiation strategy selection". This task is performed directly by the experimenter. The protocol is a bilateral negotiation protocol. The negotiation process also involves the five tasks just specified. We consider the following:

- the first agent to submit a proposal is decided by coin-tossing;
- the acceptability of a proposal is determined by a *negotiation threshold* – $ag_j\in Ag$ accepts a proposal $prop_i^{tn}$, submitted by ag_i at an instant t_n, when the difference between the benefit provided by the proposal $prop_j^{tn+1}$ that ag_j is ready to send at t_{n+1} is lower than or equal to the negotiation threshold of ag_j;
- the agents are allowed to exchange only a maximum number of proposals max_{prop} – failure to reach agreement after max_{prop} proposals results in a deadlock.

The strategies and tactics are shown in Fig. 1. The first part of the figure presents the three strategies used by both the seller and the buyer. The last part of the figure shows the five strategies used only by the buyer.

The Environment. The environment contains information about prior negotiations and market characteristics. This information is grouped into four parameters: $bfpr_X$ (base fair market value for price), $bfdp_X$ (base fair market value for down payment), $bfft_X$ (base fair market value for financing terms), and $bfdd_X$ (base fair market value for delivery date). We consider the following: (i) the base fair market values are used to compute perceived market values, and (ii) the values to offer in the opening proposal are computed from the perceived market values.

Agent	Strategy Class	Strategy Key	Opening Negotiation Tactic	Concession Tactics
Seller and Buyer	Starting reasonable and conceding moderately	SRMDT	Starting realistic	Moderate
	Low priority concession making	SRMLP	Starting realistic	Tough; Moderate; Soft
	Modified logrolling	SRMML	Starting realistic	Moderate; Soft
Buyer (only)	Starting reasonable and conceding slowly	SRSLW	Starting realistic	Tough
	Starting reasonable and conceding rapidly	SRRPD	Starting realistic	Soft
	Low priority concession making	SRSLP	Starting realistic	Tough; Soft
	Low priority concession making	SRRLP	Starting realistic	Tough; Soft
	Modified logrolling	SRSML	Starting realistic	Tough; Soft

Figure 1. Negotiation strategies and tactics.

Experimental Hypotheses

The first two hypotheses are based on the conclusions just presented. The last hypothesis is related to the process of negotiation. The hypotheses are stated as follows:

Hypothesis 1: The strategy SRMLP leads, on average, to agreements than provide higher joint benefit than the strategy SRMDT;

Hypothesis 2: The strategy SRMML leads, on average, to agreements than provide higher joint benefit than the strategy SRMDT;

Hypothesis 3: The strategy SRMLP leads, on average, to slower agreements than the strategy SRMDT.

The Experimental Method

The experimental method is controlled experimentation. A description of the experimental parameters, the independent variable, the dependent variables, and the experimental procedure follows.

Experimental Parameters. The base fair market values for price, down payment, financing terms and delivery date are set to 500, 125, 180 and 90, respectively. The perceived market values are generated by randomly choosing a value within 10% of the bases. The values to offer in the opening proposal are computed from the perceived market values. The negotiation threshold is set to 0.0 and the maximum number of proposals to 10.

The Independent Variable. The independent variable is the strategy of the seller. This variable has three levels, namely the three first strategies presented in Fig. 1.

The Dependent Variables. The first dependent variable is the joint benefit provided by the final agreement, i.e., the sum of the two agents' benefits in the final agreement. Consider that the agents agree on a proposal *prop* (specifying price *pr*, down payment *pa*, financing terms *fn* and delivery date *dt*). The benefit of each agent ag_i for price *pr* is given by the following function:

$$Vpr_i = \left| \frac{pr - lim_i}{lim_i - lim_j} \right|$$

where lim_i and lim_j are the limits of ag_i and ag_j for price, respectively. The benefits for *pa*, *fn* and *dt* are given by similar functions. The benefit for *prop* is given by an additive scoring function (Raiffa 1982).

The second dependent variable is the time spent in negotiation. This variable is measured in terms of the total number of offers exchanged by the agents until either they found an agreement or reach a deadlock. If no deal is made in a particular negotiation, then this variable is set to max_{prop}.

The Experimental Procedure. The experiment involves three groups of trials. For each group of trials, the experimenter manipulates the independent variable, i.e., assigns a strategy to the seller agent. For each trial in each group, the experimenter: (i) randomly determines the agent that starts the bidding process, and (ii) randomly determines a strategy for the buyer agent. The experimenter then allows the agents to negotiate using the strategies and measures the dependent variables.

Experimental Results

The experiment was conducted on a personal computer using Visual C^{++}. For each of the 3 groups, we conducted 30 trials. A pretest was performed to establish how many trials were needed to obtain significant averages on the measures taken. The results are shown in Fig. 2.

The main response measure was the sum of the two negotiator's benefits in the final agreement. It was predicted that the strategies SRMLP and SRMML

Group	Seller's Strategy	Seller's Benefit (mean)	Buyer's Benefit (mean)	Joint Benefit (mean)	Number of Proposals (mean)
$group_1$	*SRMDT*	0.536	0.462	0.999*,†	7.366**
$group_2$	*SRMLP*	0.615	0.449	1.064*	8.200**
$group_3$	*SRMML*	0.511	0.561	1.073†	6.533

*(F=1996.07, p<0.01); †(F=1844.40, p<0.01); **(F=5.71, p<0.025)

Figure 2. Experimental results.

[illegible] superior outcomes. The experimental results [illegible] that the strategy SRMLP resulted in significantly higher joint benefits when compared to the joint benefits resulting from the strategy SRMDT (F=996.07, p<0.01). The same is true for the strategies SRMML and SRMDT (F=1844.40, p<0.01). Hypothesis [illegible] and hypothesis 2 are supported.

The number of proposals exchanged by the agents was [illegible] recorded. The prediction was that the strategy SRMLP produced slower agreements. The results indicate that this prediction was confirmed. The strategy SRMLP resulted in significantly more proposals than the strategy SRMDT (F=5.71, p<0.025). Hypothesis 3 is supported.

RELATED WORK

The design of autonomous agents with negotiation competence has been investigated from both a theoretical and a practical perspective.

Researchers following the theoretical perspective attempt mainly to develop formal models. Some researchers define the modalities of the mental state of the agents, develop a logical model of individual behavior, and then use the model as a basis for the development of a formal model of negotiation or argumentation (e.g. Kraus et al. 1998). However, most researchers are neutral with respect to the modalities of the mental state and just develop formal models of negotiation. These models are often based on game-theoretic techniques (e.g. Fatima et al. 2004; Kraus 2001). Generally speaking, most theoretical models are rich but restrictive. They make assumptions that severely limit their applicability to solve real problems.

Researchers following the practical perspective attempt mainly to develop computational models, *i.e.*, models defining the key data structures of the agents and the processes operating on these structures. Some researchers start with a model of individual behavior, develop or adopt a negotiation model, and then integrate both models (e.g., Muller 1996). Again, most researchers prefer to be neutral about the model of individual behavior and just develop negotiation models (e.g. Faratin et al. 2002). Broadly speaking, most computational models are rich but based on ad hoc principles. They lack a rigorous theoretical grounding. Despite these weaknesses, some researchers, including the authors, believe that it is necessary to develop computational models in order to successfully use agents in real-world applications. Accordingly, this paper presented a computational model of negotiation.

As noted, most researchers have paid little attention to the problem of integrating models of individual behavior with negotiation models. However, it is one of the earliest lessons of computer science that independently developed components resist subsequent integration in a smoothly functioning whole. Components need to be designed for integration right from the start. Accordingly, this paper presented the key features of a model that accounts for a tight integration of the individual capability of planning and the social capability of negotiation.

CONCLUSION

This paper presented the key features of a negotiation model for autonomous agents. The model handles multi-party and multi-issue negotiation, acknowledges the role of conflict as a driving force of negotiation, formalizes a set of human negotiation procedures from management and social psychology and combines them with AI techniques, and accounts for a tight integration of individual and social behavior.

This paper also described an experiment performed to empirically evaluate a version of the model that handles two-party, multi-issue negotiation. The experimental results showed that: (i) the "low priority concession making" and "modified logrolling" strategies lead, on average, to superior outcomes, and (ii) the "low priority concession making" strategies lead, on average, to slower agreements. The results confirmed two conclusions about human negotiation. Our aim for the future is: (i) to extend the negotiation model, and (ii) to continue the experimental validation of the model.

REFERENCES

Carnevale P. and D. Pruitt. 1992. "Negotiation and Mediation", In: *Annual Review of Psychology*. Annual Reviews Inc., 43, 531-581.

Faratin, P.; C. Sierra and N. Jennings. 2002. "Using similarity criteria to make issue trade-offs in automated negotiations". *Artificial Intelligence*, 142: 205-237.

Fatima, S.; M. Wooldridge and N. Jennings. 2004. "An agenda-based framework for multi-issue negotiation". *Artificial Intelligence*, 152: 1-45.

Kraus, S.; K. Sycara and A. Evenchick. 1998. "Reaching Agreements Through Argumentation: a Logical Model and Implementation". *Artificial Intelligence*, 104 (1-2): 1-69.

Kraus S. 2001. *Strategic Negotiation in Multi-Agent Environments*. The MIT Press, Cambridge, MA.

Lewicki, R.; B. Barry, D. Saunders and J. Minton. 2003. *Negotiation*. 4th edition, McGraw Hill.

Lopes, F.; N. Mamede, A. Q. Novais and H. Coelho. 2002. "A Negotiation Model for Autonomous Computational Agents: Formal Description and Empirical Evaluation". *Journal of Intelligent & Fuzzy Systems*, 12: 195-212.

Lopes, F.; N. Mamede, A. Q. Novais and H. Coelho. 2004. "Negotiation Strategies for Autonomous Computational Agents". In *ECAI-04*, IOS Press, 38-42.

Muller J. 1996. *The Design of Intelligent Agents*. Springer-Verlag, Berlin (LNAI 1177).

Pruitt D. and S. Kim. 2004. *Social Conflict: Escalation, Stalemate, and Settlement*, 3rd edition, McGraw Hill.

Raiffa H. 1982. *The Art and Science of Negotiation*. Harvard University Press, Cambridge.

AGENT-BASED PROCESS MONITORING: APPLYING EVENT MANAGEMENT PARADIGMS TO A STRATEGIC PLANNING PROCESS

Stefan Winkler
Roland Zimmermann
Bernd Vögele
Freimut Bodendorf
Department of Information Systems
University of Erlangen-Nuremberg
Lange Gasse 20, 90403 Nuremberg, Germany
{Stefan.Winkler; Roland.Zimmermann; Freimut.Bodendorf}@wiso.uni-erlangen.de

KEYWORDS

Process Monitoring, Agents, Event Management

ABSTRACT

The monitoring and coordination of planning processes requires a very flexible support with information systems. Due to the high amount of activities conducted in parallel as well as numerous interdependencies between inputs and outputs of every process step, present coordination concepts result in information deficits during process execution. In cooperation with an industry partner an agent-based information logistics concept is developed to reduce these deficits. The approach builds upon event management paradigms which have shown significant improvements in supply chain management. A prototype is realized and serves as the basis for evaluation.

PROBLEM AND OBJECTIVE

Sales planning processes in large companies encompass development of strategic guidelines, operational planning of sales quantity, turnover to be achieved, and future sales activities. All planning results have to be agreed upon by different organizational units and must then be aggregated before being approved by top management.

Further analysis considers specific sales planning processes of an industry partner. These processes are conducted according to a predefined process model by multiple actors in many different countries. The main problem is coordination of these activities which are often dependent upon each other but are not well synchronized. In many cases results of a preceding activity are required as a precondition for some succeeding activity. These results arrive too late and information on available results is not communicated. Thus, an information deficit exists among the actors who are responsible for the planning processes. A high degree of complexity of the coordination problem results from the fundamental characteristics of the processes:

- A variety of dependencies exists between individual activities and activity results. To initiate a single activity a set of inputs such as e.g. local sales quantities are required. These figures in turn are the output of preceding process steps.
- Outputs (and partial results) of certain activities within the planning process differ considerably ranging from only slightly structured (e.g. texts or guidelines) to highly structured outputs (e.g. tables or spreadsheets).
- The planning process frequently consists of decision making activities rather than purely administrative functions. Employees have to make decisions based on results of previous process steps as well as additional information available to them. Resulting planning values are communicated to other actors.
- Temporal dependencies between sub-processes are of considerable significance. Changes to the time schedule are common during the course of the planning process.

Shortcomings in the Area of Information Logistics

Current process management at the industry partner is restricted to modelling processes and a large number of the resulting documents with a graphical notation. The process models represent the process steps in a clearly structured manner and are available for reference purposes to all actors involved in sales planning. However, from an information logistics perspective, no coordination support is provided and the identified information deficit is not tackled. Some characteristic shortcomings are:

- Existing enterprise applications are characterized by low levels of data and functional integration.

– A global data model as well as management of access rights and different document versions has not (yet) been implemented with respect to relevant planning documents.

– Existing coordination mechanisms take insufficient account of temporal changes and dependencies during the planning process.

– Automated feedback among actors in the planning process is not possible. All communication regarding process results has to be initiated manually.

Consequently, it is assumed that parties involved in the planning process do not have all information relevant to them and known to the enterprise at the time of the decision. Obtaining this information is associated with higher costs than necessary. Finally, effects of disruptions (e.g. delays) on subsequent process steps are not apparent to affected actors.

Requirements and Objectives

The overall objective of the suggested concept is to bridge the gap between simple process models and automated support of decentralized execution of the processes in question. As described above, process models are usually generated during the first phase of process management using fully developed modelling tools. To support the transition from process models to a distributed support system with a focus on the requirements of a flexible information logistics architecture, agent technology is employed.

To relieve actors of the burden of obtaining relevant information as well as getting the current status of related process steps from other actors, software agents are introduced which are able to act autonomously. According to the different needs of actors in the process steps, software agents proactively request information from various resources and communicate their status to the agents of all related activities. Users are notified every time the status of related process activities has changed (e.g. all input data is available) or certain deadlines during the course of the process are being exceeded.

To ensure usability of the agent-based system the approach has to cope with the various restrictions regarding the highly distributed execution of the planning process. Synchronization of tasks realized in parallel, information exchange using different means of representation as well as access to heterogeneous information sources such as business applications or database systems are just a few examples of further requirements of the process monitoring system.

Finally, changes in the planning process model (e.g. as a result of business process reengineering) shall be reflected in the agent-based information logistics system automatically to support integrated process monitoring. Therefore, all participating software agents implement a configuration component which allows them to adapt easily to changes concerning the process flow or the access to different information resources.

RELATED WORK

Common standard software products for the execution of business processes are Workflow Management Systems (WfMS). These systems are used to coordinate business processes. The integration of process modeling and workflow instantiation through translation of process models into workflow configurations is realized by Business Process Management Systems (BPMS). However, WfMS and BPMS are primarily applied to highly structured and frequently executed processes. In general, neither of these systems provides decision relevant information to the user in a proactive manner. This deficit is especially evident, if information is distributed among multiple applications and has to be collected and evaluated individually.

Service oriented support systems based on web services are increasingly important (e.g. BEXEE, http://bexee.sourceforge.net/). In particular, Serviceoriented Architectures (SoA) offer flexible integration of different information sources as well as orchestration of web services which is for instance based on BPEL4WS or WS-BPEL. However, web services still lack the proactivity and autonomy necessary to provide information logistics services in processes that are not highly structured.

Software agents offer the ability to imitate human coordination and cooperation mechanisms due to their properties such as autonomy, reactivity, proactivity and social behavior (Jennings et al. 1998). Hence, software agents can offer information logistics support similar to that which human actors can provide.

Agent software has reached a level of sophistication that allows it to be used to design and implement industrial strength applications. Within a prototype at DaimlerChrysler production of cylinders is coordinated with the aid of agent-based negotiation mechanisms (Bussmann and Schild 2000). Agents are used in electronic market places where they buy and sell goods. One provider of such solutions is e.g. Whitestein Technologies (http://www.whitestein.com). Further applications in particular for C2C markets are presented e.g. by Eymann (Eymann 2003).

Information logistics solutions are provided within the scope of Supply Chain Event Management (SCEM). Well-engineered concepts exist for inter-organizational monitoring of orders (Bodendorf and Zimmermann 2005, Zimmermann et al. 2004). To improve acceptance of agent based applications in the industrial context, different approaches for adjustable autonomy are discussed by e.g. Hexmoore (Hexmoore et al. 2003). Further research relevant to the information logistics problem has been conducted in the fields of robustness of agent systems as well as security and cryptography.

AGENT-BASED CONCEPT

Suitability

In general agent systems are particularly suitable for complex, decentralized systems because they effectively support the principles of decomposition, abstraction and flexible organization which are required for such a system (Jennings 2004).

The advantages of using agent technology as opposed to other options in the context of the process management task are due to agent characteristics:

- *Proactivity*: Agents actively request information from other agents on demand and communicate with their human users.
- *Reactivity*: Status inquiries are answered immediately whereas new status information is selectively broadcasted.
- *Social ability*: Agents are able to communicate with other agents and human users.
- *Autonomy*: Warnings are autonomously escalated to the responsible agents or users, if an input factor reaches a critical value.

The agent based approach adopted here supports a decentralized coordination. Agents are (partly) autonomous software components that represent autonomous units such as decision makers or parts of an organization responsible for decision making.

Agent Types

In the context of process monitoring agents support process activities as well as business roles. The system design is based on the assumption that there can be several decisions for each role (n:1-Relationship). But there should be only one person who is responsible for the decision (1:1-Relationship). This restriction also reflects the target state of the process. Hierarchical relationships between decisions ("aggregating decisions") are possible as well. This allows the system to support multiple process levels in hierarchical process models.

A so called activity agent is assigned to every activity which is relevant for a decision. This agent permanently monitors the state of the activity. The user (role) accesses the activity agents using another type of agent, the so called visualization agent. Possible access methods are status requests or the input of parameter values. The visualization agent is responsible for the representation of several decisions of a specific role. If necessary the representation is customized according to the role of the current user. The configuration is shown in Figure 1.

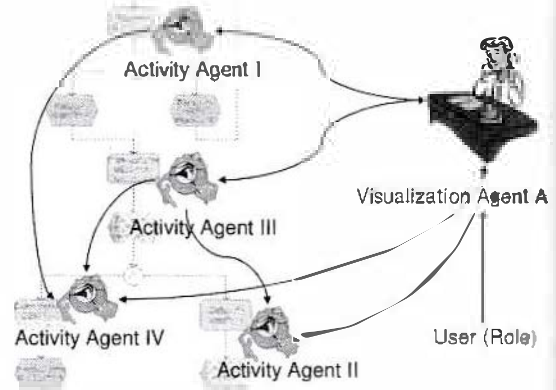


Figure 1: Agent society

Due to the characteristics of activity agents and visualization agents, both types can be easily adapted to a certain process model. To reduce implementation and maintenance of the activity agents their functionality is limited to support only one decision. Moreover, central management of all activity agents for one role is easier to handle and ensures a higher degree of availability.

The drawback of the use of different types of agents is the increasing complexity of the overall system which results from the extensive agent communication. Furthermore, companywide use would require an authorization management which accomplishes mapping between roles and associated activities.

Agent Behavior

The individual parties involved largely decide independently and with the aid of highly heterogeneous application systems (e.g. MS-Office products, databases). Results of a decision are communicated and visualized by agents. Consequently, the level and quality of information available to all actors involved in a decision are improved.

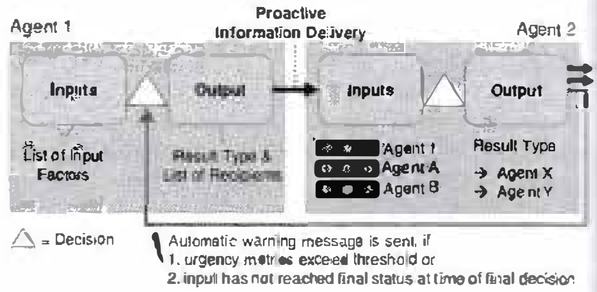


Figure 2: Agent-based information logistics

The basic principle of the information logistics concept used is the proactive delivery of information to and from the software agents. Changes in status and interim results are communicated by the responsible agent to all agents who are defined as parties interested in the [illegible]

These in turn send warning messages if temporal or logical restrictions are violated (see Figure 2). The user interface is continuously updated with the current status of a decision.

Visualization uses a traffic light metaphor. If all results required as a precondition for an activity are available in time, the traffic light is set to green. If the due date of an activity is reached and e.g. a single input is still missing, the traffic light is set to yellow. Finally, a red traffic light notifies the user that a given deadline has been exceeded and necessary results of an activity are still missing.

Rule-based Communication

The agent system communicates with two different types of recipients according to a set of communication rules. On the one hand, there are agents and other software components. On the other hand, agents are able to communicate with human actors, the users. Hence, two categories of communication rules are supported:

- *Communication rules between agents:* These rules specify when an agent *A* has to send a warning or a message to an agent *B*. Furthermore, the type of the message is determined.
- *Notification rules between agent and user:* The notification rules are used by an agent *A* to decide under which circumstances the user has to be notified.

In the case of communication between agents the current state of the result is exchanged after acknowledgement by the user. Notification of a user is triggered by values of different variables. The *state* of a result type indicates its condition for a given user at the current time. The *urgency* is a numeric value that shows how urgent an input for a given decision needs to be delivered. Finally the *time* represents the system time of the agent system. In relation to a deadline for an activity, time influences *urgency* of follow-up messages.

On the communication level between software agents and human users notification rules determine the proactive behavior of the agent system. According to these rules the agents continuously evaluate the variables as described above and inform the user if e.g. a given threshold is exceeded. A configuration interface allows the user to adjust both types of communication rules.

Architectural Approach

The implementation approach is based on the *Presentation Abstraction Control* pattern (PAC pattern, see Buschmann 2002). With respect to the requirements of process monitoring and given standards (e.g. FIPA - Foundation for Intelligent Physical Agents, http://www.fipa.org) the architecture will be adapted as appropriate.

Two distinct types of agents and a layered agent model (PAC-Model) result hereby. Decisions which are subordinated in terms of hierarchy are represented as a specific input type and do not require a separate agent type. This serves to reduce the complexity of the system.

The separation of concerns (Dijkstra 1976) inside the system occurs between the activity agents and the visualization agents (see Agent Types). Each activity agent refers to exactly one decision making activity and examines its status. A visualization agent by contrast refers to a certain role and shows the user interfaces of several activity agents. Agents are divided into three separate layers respectively (see Figure 3):

- The *presentation layer* manages the user interaction.
- The *control layer* encapsulates the communication and business logic.
- The *abstraction layer* separates the data access.

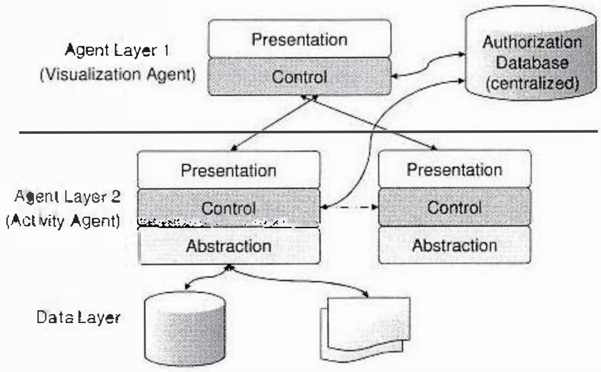


Figure 3: Agent layer model

Agent Interactions

The FIPA reference model specifies interaction patterns which for example enable FIPA-compliant agents to negotiate (see FIPA "contract net" as an example). In addition to such complex and predefined structures new interaction protocols can be defined on the basis of a request-inform-mechanism or other sub-protocols.

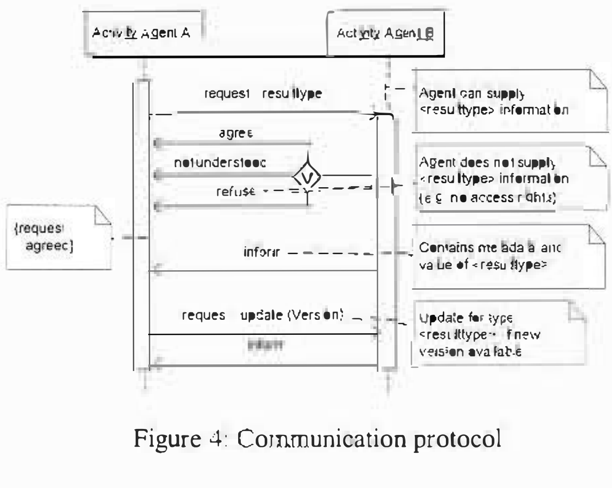


Figure 4: Communication protocol

Figure 4 shows the communication protocol of the agent system used by two agents. An activity agent *A* registers himself with an activity agent *B* to be kept informed about the state of the result type of agent *B*. The graphical notation represents a sequence diagram based on Agent UML (Odell et al. 2001), an extension of the widely used Unified Modeling Language (UML).

PROTOTYPE

The user interface represents a role's local view on all activities the role is responsible for. Each activity is represented by a tab. The basic layout of the user interface corresponds to the direction of the process flow: The relevant inputs are shown on the left. All input fields and additional information for the active decision is placed in the middle. Finally, the recipients of the decision result are shown on the right (see Figure 5). The different types of variables used for inputs or receivers are separated within different branches in the tree representation.

An example process (see Figure 6) is used to present the three main objectives and functionalities of the system: providing the user with easy access to relevant process information, communicating status changes as well as sending warning and feedback messages. Customization functions adapt the behavior with regard to the rules for communication of events and notification of users. Since many similar process instances are usually running in parallel, a time simulation component enables the system to simulate the effects of a particular setup in accelerated time or assess the projected urgency values of missing results at a given date. Therefore, this component may be used to find a suitable setup for communication rules to e.g. prevent information overload, or to help analyze the planning process as such.

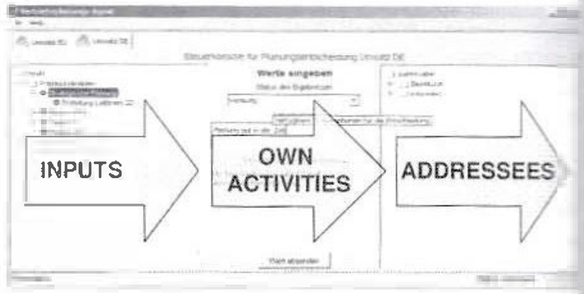


Figure 5: User guidance through GUI

The main window and several information and configuration dialogs offer direct access to the stakeholders. This includes technical process information from the company's process model documentation and contact information from the enterprise user directory. Additionally, the user can administer and update information regarding his own activity (e.g. status or remarks).

With regard to the urgency metric, which is calculated according to a predefined formula, the user can determine the threshold at which events are generated and propagated to the addressees' agents (i.e. customers) or passed on to the input-related agents (i.e. suppliers). During the course of the process the user is presented with a permanently updated view of all parties involved in or affected by each activity's result. This level of transparency is not limited to the immediate predecessors and successors in the process chain. All those other activities from which the user's activity has received warning messages can be viewed as well, thereby enabling the user to locate the origin and development of a disruption while filtering out those parts of the overall process that are irrelevant to him.

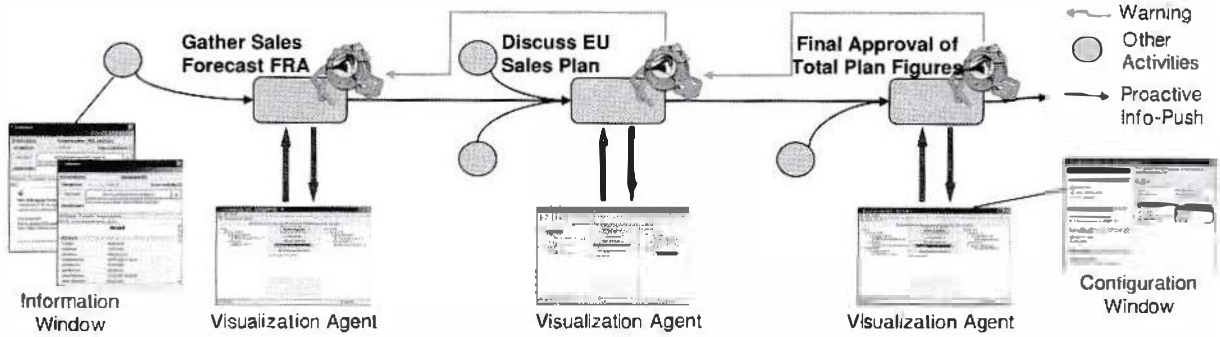


Figure 6: Example process in showcase

EVALUATION

In addition to the general suitability of the agent-based concept, factors concerning the existing system environment, the robustness of the infrastructure and the available know-how are of importance for the operative use in a company. Integration with process support software is generally conceivable, in particular the combination of agent systems with workflow management systems is proposed (Schönfeldt 2001). If strict company-wide IT-guidelines regulate the use of particular products, two approaches are viable: Either try to connect the prototype presented herein with the corre-

...ending products via plug-in interfaces. Or alternatively implement the concept as far as possible using the means of existing products, possibly with a reduction in functionality.

In reviewing agent systems a distinction must be drawn between the technology and the paradigms of the agent based approach. While the use of a dedicated agent platform, as in this work, has its advantages, it is by no means a prerequisite for the development of software agents. In this regard, it should be examined whether the portal or middleware system already introduced in the company can be extended with reasonable efforts to include those features the presented agent system provides.

Regarding the fulfillment of ergonomic requirements the prototype was assessed using the criteria specified in the ISO 9241/10 standard (Prümper and Anft 1993):

- *Suitability for the task*: The prototype was specifically developed with the objectives of the use thereof in mind. Manual inputs are further reduced by virtue of the fact that it is connected to existing systems.

- *Self descriptiveness*: Technical terms should be known to the user (this can be assumed to be the case in the target group of expert users). A situation specific help system is still to be implemented.

- *Controllability*: A flexible method of working is possible because the dialogues are not modal and values are updated proactively. However, popup-warnings have to be confirmed.

- *Conformity with user expectations*: The user receives constant reports via the protocol function; due to the principle of agent systems, the response times of other agents cannot be predicted precisely.

- *Error tolerance*: The protocols and dialogues issue error reports, solution tips are still largely absent.

- *Suitability for individualization*: A high level of adaptability is achieved by using a rule-based system. A beginner mode has not been realized (and would only be of limited benefit because the software is used by experts).

- *Suitability for learning*: If the concept is known and understood, the functions are accessible via a graphic user interface and motivate the user to fully avail of the level of functionality available.

CONCLUSIONS

The proposed agent-based approach meets the demands of process monitoring. A showcase for a realistic part of the planning process of the industrial partner is realized and documented. Improved coordination reduces transaction costs and increases the availability of decision-supporting information for human actors. The information deficit is reduced. Furthermore, an initial assessment of the prototype from a user perspective based on an ISO norm questionnaire indicates that users accept this form of automated decision support. Further work focuses on integration of the prototype in a productive environment which requires certain adjustments of the underlying agent-based infrastructure. Finally, development of process monitoring is extended to fulfill the needs of different types of processes in varying application domains.

REFERENCES

Bodendorf, F. and Zimmermann, R. 2005. "Proactive Supply Chain Event Management with Agent Technology". In *International Journal of Electronic Commerce* 9, Inderscience Enterprises Ltd., 57-90.

Buschmann, F. 2002. *Pattern-Oriented Software Architecture: A System of Patterns*. 2nd Ed. Wiley, Indianapolis.

Bussmann, S. and Schild, K. 2000. "Self-organizing manufacturing control: an industrial application of agent technology". In *Proceedings 4th Int. Conf. on Multiagent Systems*, ICMAS'2000. Boston, 87-94.

Dijkstra, E.W. 1976. *A Discipline of programming*. Prentice Hall, Englewood Cliffs, NJ.

Eymann, T. 2003. *Digitale Geschäftsagenten*. Springer, Berlin.

Hexmoore, H.; Castelfranchi, C.; and Falcone, R. 2003. *Agent Autonomy*, Kluwer Academic Publishers.

Jennings, N.; Sycara, K.; and Wooldrige, M. 1998. "A Roadmap of Agent Research and Development". In *Autonomous Agents and Multi-Agent Systems* 1, Kluwer Academic Publishers, 7-38.

Jennings, N. 2004. "An Agent-based Approach for Building Complex Software Systems". In *Communications of the ACM* No. 4, Vol. 44 (2004), 35-41.

Odell, J.; Parunak, H.; and Bauer, B. 2001. "Representing Agent Interaction Protocols in UML". In *UML, Agent-Oriented Software Engineering*. P. Ciancarini and M. Woolridge (Eds.). Springer, Berlin, 121-140.

Prümper, J. and Anft, M. 1993. *Beurteilung von Software auf Grundlage der Internationalen Ergonomie-Norm ISO 9241/10*. FHTW Berlin.

Schönfeldt, B.F. 2001. *Einsatz intelligenter Agenten zur Entscheidungsunterstützung in prozessorientierten Unternehmen*. Dissertation, Universität Mannheim.

Zimmermann, R.; Butscher, R.; Bodendorf, F.; Huber, A.; and Görz, G. 2004. "Unternehmensübergreifendes Event Management mit Softwareagenten". In *Supply Chain Management in der Automobil- und Zulieferindustrie*. W. Dangelmaier; D. Kaschula; and J. Neumann (Eds.), ALB-HNI, Paderborn, 79-88.

ASSESSMENT OF SUPPLY CHAIN EVENT MANAGEMENT DATA BY AN AGENT-BASED SYSTEM WITH FUZZY LOGIC

Roland Zimmermann
Stefan Winkler
Clemens Meyreiss
Freimut Bodendorf
Department of Information Systems
University of Erlangen-Nuremberg
Lange Gasse 20, 90403 Nuremberg, Germany
{Roland.Zimmermann; Stefan.Winkler; Freimut.Bodendorf}@wiso.uni-erlangen.de

KEYWORDS

Supply Chain Management, Agents, Fuzzy Logic

ABSTRACT

Software agents which gather event-related information for *Supply Chain Event Management (SCEM)* purposes are confronted with a complex task: analysis of diverse SCEM data. Fuzzy Logic provides mechanisms for heuristic human-like assessments of these data types. A concept for integration of Fuzzy Logic in software agents is presented which is implemented in a prototype. Evaluation of experiments indicates high quality of an agent's Fuzzy Logic analysis results compared to a human actor's heuristic assessments of the same data.

PROBLEM

Supply Chain Event Management promises to identify and correct distuptive events and malfunctions in operational supply chain processes by providing event-related information to decision makers in a timely fashion. An agent-based concept for event management in multi-level supply chains is described in (Bodendorf et al. 2005). It is shown in experiments and in an industrial showcase that monetary benefits of agent-based SCEM are significant (Zimmermann et al. 2005). This concept includes software agents which proactively gather data on orders and related suborders that have been placed with suppliers. Typical data types collected by an agent are planned and estimated dates of delivery for orders and information on disruptive events. One of the major tasks for the software agents is to analyze and interpret gathered data automatically. They decide upon their own assessments whether alerts to actors in a supply chain have to be generated.

Any analysis of event-related data is influenced by developments in fulfillment processes of monitored orders (e.g. production, transportation). These processes are executed by a large variety of actors and resources which influence each other directly or - even more often - indirectly. An example is a disruptive event "traffic jam" which affects transportation processes. It is caused by a multitude of actors - all vehicle drivers within the congestion - and additional factors such as e.g. weather conditions. Consequently, its duration cannot be accurately forecasted with reasonable efforts. Moreover, effects on orders transported by a certain truck which is stuck in the traffic jam cannot be predicted for certain either, e.g. due to unforeseeable reactions of the truck driver. Hence, each data gathering agent is confronted with various types of data and disruptive events in a multitude of environmental settings. It is not possible to model all influencing factors that would be required to exactly forecast consequences of a disruptive event for an order's future fulfillment. Nevertheless, a human actor is able to gain important insights into an order's status, if data on disruptive events and process performance measurements are available: He generates heuristic interpretations for different aspects of an order's situation. Software agents imitate this heuristic approach.

DATA INTERPRETATION WITH FUZZY LOGIC

Simple calculations (e.g. weighted averages of input data) or simple decision rules (If...Then...Else) are not applicable for a heuristic interpretation which has to be similar to a human actor. Especially the vagueness of implications associated with gathered event management data has to be represented quite like a human actor would assess the situation. For this reason an approach based on Fuzzy Logic is chosen. In contrast to other methodologies, Fuzzy Logic is able to reason with perceptions (Zadeh 1999). Zadeh argues that a perception is a fuzzy evaluation of a concept such as time, distance, weight, likelihood, or truth. An example is "warm" as a perception of temperature. It is opposed to the concept of a measurement which is represented by an exact value (e.g. a temperature of 25.6° Celsius). SCEM data types which are the input to an agent's analysis process are considered to be measurements. An assessment of a situation represented by these measurements has to consider both, the perceptions a human actor would experience regarding these measurements and the reasoning he would apply based on these perions. This is achieved by using **Fuzzy Logic** - a combination of fuzzy perceptions and mathematically founded logic (Friedrich 1997, pp. 161).

AGENT SOCIETY

Overview

To realize the SCEM concept within a supply chain, each supply chain partner provides one agent society with a discourse and a coordination agent, as well as various surveillance and wrapper agents (see Figure 1) (Bodendorf et al. 2005). A single coordination agent in each enterprise assures that initialization of monitoring efforts as well as management of external status requests and alerts is handled consistently within an enterprise. The coordination agent also provides an overview of all monitored orders of an enterprise and serves as a management cockpit for event management activities.

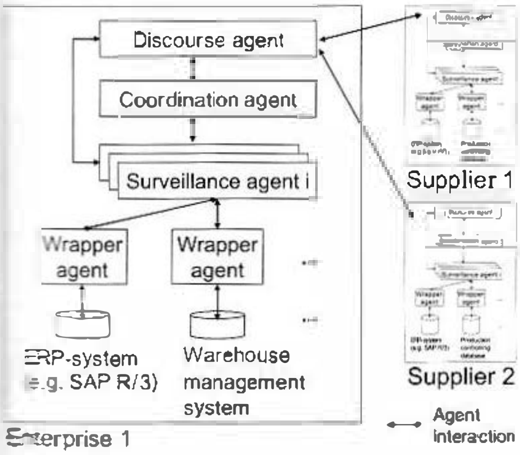


Figure 1. Agent society

For each monitored order of an enterprise a dedicated surveillance agent is triggered by the coordination agent. Thus, the data gathering and analysis functions are encapsulated in dedicated surveillance agents while the coordination agent decides on generation of alerts. Wrapper agents provide a standard interface to internal data sources for surveillance agents. Fuzzy Logic mechanisms are used in both, surveillance and coordination agents.

Surveillance Agents

Different types of SCEM status data are used to calculate deviations from plans, e.g. delays, incomplete quantities or quality measures derived from quality assessments. A human actor who assesses an order's situation considers various such indicators and generates an overall assessment of the order's status. Similarly, a surveillance agent integrates a variety of these inputs to form an aggregate assessment which is termed the *Aggregated Order Status (AOS)*. Calculation of an *AOS* by a specific enterprise is influenced by its strategic goals. For instance, a differentiation strategy based on very high product quality requires to rate quality misses of suppliers higher than delays. A surveillance agent considers these strategic implications for its assessment.

Disruptive events are identified by a surveillance agent during fulfillment of either its monitored order or one of its respective suborders. These events have different effects depending on the time of their identification relative to the remaining fulfillment time of an affected order. The same event (e.g. a machine breakdown) tends to have more serious consequences, if it takes place close to the end of a production process and thus an order's planned fulfillment date: The remaining reaction time is reduced, compared to an earlier identification of the same type of event, and associated follow-up costs rise. Hence, a surveillance agent considers the planned timeline of a process and assesses the severity of an event based on the current fulfillment situation of an affected order. This results in an order specific measurement of a disruptive event's severity, termed the *Endogenous Disruptive Event Severity (EnDS)*.

Coordination Agent

The coordination agent decides whether to generate any alert for a certain order. Data considered in this decision encompasses results of the Fuzzy Logic data analysis conducted by a surveillance agent and further information such as the priority of an order. This mechanism employs a two-step Fuzzy Logic process. It results in an abstract metric value termed *Alert Index (AI)* that is used in subsequent steps by the coordination agent to decide on generation of an alert and to determine recipients and media types.

FUZZY LOGIC ANALYSIS

Aggregated Order Status

SCEM data for a variety of status assessments regarding a specific process (e.g. production) is gathered by a surveillance agent from internal data sources and from suborder recipients. The indicators which are derived from this data are differentiated into absolute and relative indicators: Depending on what types of indicators (e.g. time vs. quality) are to be considered in the *AOS* and on the characteristics of monitored orders (respectively their suborders), either absolute or relative indicators are better suited. For instance, if one suborder has a planned fulfillment duration of two weeks while another suborder of the same order has only two days, a relative indicator "% delay" is not suitable: A 10% delay of the first suborder (~1.5 days late) will affect the superordinate order much more than a 10% delay of the second suborder which is then only about five hours late. However, relative indicators are often used in quality measurements, e.g. a percentage of defect parts in a delivery. These indicators facilitate comparison of different environmental situations (e.g. deliveries of different size).

Any indicator which is used in the Fuzzy Logic analysis process of the surveillance agent is fuzzified in a first step. For each indicator a linguistic variable with different fuzzy variables is defined. An applicable membership function for fuzzy sets in this domain is the trapezoid function. It is suitable for indicators that can be derived from status data types since a human actor typically perceives a deviation within a certain range as high or critical with a value of one (e.g. critical=1). Only the transition to the next fuzzy set (e.g. high to very high) is valued in between one and zero.

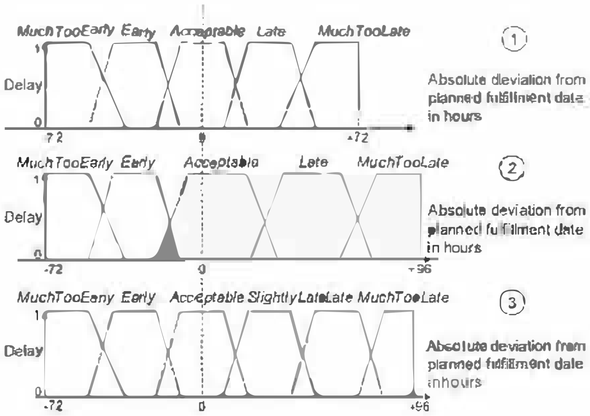


Figure 2. Alternative fuzzy sets

In the example in Figure 2 the linguistic variable *Delay* is defined based on five fuzzy variables within a range of 72 hours before and after the planned fulfillment date of an order (1). Depending on the strategic goals and the specific industry of a supply chain partner, different definitions of delays can be configured. In Figure 2 two other possible definitions are depicted that spread three fuzzy sets to allow for longer delays (2) and that add a sixth fuzzy set to further differentiate delays (3).

To assess fuzzified input values a fuzzy rule set is required which allows creation of an *AOS*. The *AOS* is standardized in the range between zero and one. It is defined as a linguistic variable with fuzzy sets *VeryHigh* for fulfillment that is as planned and *VeryLow*, if large problems are identified. Three intermediate fuzzy sets complete this linguistic variable. For two basic input values - absolute delay of an order (*ProcessTimeAbs*) and absolute deviation from ordered quantity (*ProcessQuantAbs*) a graphical representation of a possible rule sets is given (see Figure 3). The rule set reflects a typical just-in-time strategy of a manufacturer which depends on timely deliveries from its suppliers and has (nearly) no capacities for safety stocks. Both, late or incomplete deliveries result in high follow-up costs for the manufacturer, because his production lines are halted soon, if input material is not delivered continuously. Thus, every kind of late delivery and every type of incomplete delivery is rated very critical and results in a low *AOS*.

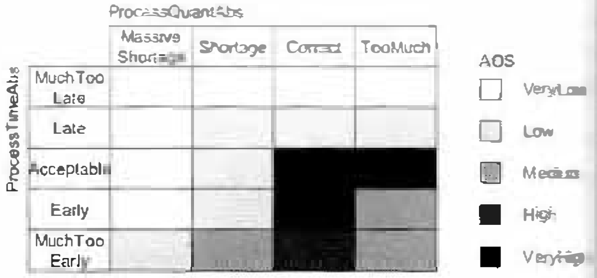


Figure 3. AOS - definition of fuzzy rules

By comparing the fuzzy rule base to the percepti[...] associated with the input values, a number of evalu[...] tions is generated for each perception. These evaluatio[...] are aggregated, and a single value for the *AOS* is calc[...] lated using a defuzzification method (e.g. the center [...] gravity). This *AOS* allows to characterize a monitored order's status. For instance, a value of 0.23 with a possible interval of the *AOS* between zero and one indicates [...] relatively high current criticality of a monitored order.

A surveillance agent gathers the same data types [...] data on its monitored order as well as its related subor[...] ders. Thus, a number of data sets with similar da[...] inputs have to be aggregated and then interpreted by [...] surveillance agent. A filter is used to select the mo[...] important SCEM data inputs and forward these to [...] fuzzy analysis component of the agent. Depending [...] how this filter is configured, the agent realizes [...] individual strategy of its company. A typical strategy [...] to select worst cases for each type of indicator a[...] forward these to the Fuzzy Logic analysis.

Summarizing, the *AOS* is an individual assessment [...] a monitored order's situation which incorporates diffe[...] ent status aspects of an order and its relevant suborders. The *AOS* is calculated whenever new SCEM informa[...] tion becomes available to a surveillance agent and th[...] the *AOS* changes over time. The assessment refle[...] individual valuations and strategies that vary for eac[...] supply chain partner.

Endogenous Disruptive Event Severity

Disruptive events which are identified by a surveil[...] lance agent have to be analyzed as to their effect [...] fulfillment processes, in spite of the fact that a comple[...] model of cause-and-effect for each type of event is [...] feasible (see above). As requested an event is analyzed with respect to the planned timeline of the fulfillmen[...] processes it affects. Two input values are needed:

- An external classification of a disruptive event['s] severity is a measurement of severity which is assumed to be defined for each type of event a[...] which is derived for instance from a ranking list wi[...] associated severity values. As an example, a machine failure is rated lower than a power outage. For eac[...] event a classification value between zero and one [...] assumed which is referred to as the *Exogenous Di*[...]

ruptive Event Severity (ExDS)*. This severity is independent of the time of occurrence of an event and is fixed.

- The *Remaining Time (RT)* to a planned fulfillment date is considered under the assumption that an event has a larger negative impact on an order's fulfillment the later it occurs in a fulfillment process and the less time for reaction remains. It is defined as the difference between the planned end date of fulfillment of an order and the date of identification of an event by a surveillance agent.

Using a similar Fuzzy Logic mechanism as for the AOS, a so called *Endogenous Disruptive Event Severity (EnDS)* is determined by a surveillance agent. The *EnDS* reflects a heuristic assessment of the probability to solve the problem that is caused by an event in the remaining planned fulfillment time of an order. A high *EnDS* indicates that propagation of an event to the next supply chain level is highly likely, whereas a low *EnDS* characterizes an event that is solvable within an enterprise. Consequently, *EnDS* is used to determine whether a specific event has to be communicated by a supply chain partner in a message to its customer. Disruptive events with a low *EnDS* are not communicated, in order to avoid an information overflow on following supply chain levels through irrelevant data. Calculation of an *EnDS* for an event is only initiated once for each event identified by a surveillance agent, because its parameters (*ExDS, RT*) remain constant as long as no corrections such as a revision of *ExDS*) occur, in which case a recalculation is initiated.

Alert Index

The coordination agent determines an *Alert Index (AI)* for each order based on the analytical results provided by surveillance agents. Input values for the *AI* are e.g. AOS and the maximum *EnDS* of all new disruptive events identified in the last data gathering round by a surveillance agent. Additional data types that a company wants to consider for its alert generation (e.g. a customer's rating) are incorporated in a second step of the coordination agent's Fuzzy Logic analysis.

A two-step stacked Fuzzy Logic process is chosen to limit the complexity of the fuzzy rule sets. An example of a fuzzy rule set for the first step may represent a very cautious strategy regarding the condition of an order. Such a rule set considers severe disruptive events (*VeryHigh EnDS*) as very important and thus raises the *AI* to the highest level even if the corresponding aggregated order status *AOS* is very high. The strategy is justified under the assumption that a newly discovered severe disruptive event has not yet affected an order's status data and its negative consequences thus have not yet been measured. However, effects on status data will be reflected in future data gathering rounds while the *AI* is raised instantaneously to a very high level which permits reactions even before any negative consequences of the event are encountered.

The second Fuzzy Logic step is independent of the first step. For instance, a company may value some orders higher than others, depending on their priority. The priority of an order is a value that is determined outside a SCEM system by each supply chain partner. Important sources for definition of an order's priority are e.g. marketing and sales departments that have data and strategies in place to define order priorities. Possible input values are: sales revenues with a customer, profit margin of an order or service level agreements with customers. Ideally, a standardized value for an order's priority is provided that is for instance calculated based on a multi-dimensional scoring model.

After the second Fuzzy Logic step a defuzzification mechanism provides a metrical value between zero and one. This final *AI* is used by the coordination agent to decide on alerts. This decision includes a discrete escalation mechanism not detailed here.

PROTOTYPE

A generic prototype with all agent types is realized for conducting experiments in a laboratory environment: Each enterprise in a simulated supply chain hosts one agent society. The main focus of the implementation is on SCEM features provided by coordination and surveillance agents, whereas only basic mechanisms of discourse and wrapper agents are realized. Every agent society is realized on its own instance of the FIPA-conform JADE agent platform. Examples of visualizations for analytical results of a surveillance agent are depicted in Figure 4.

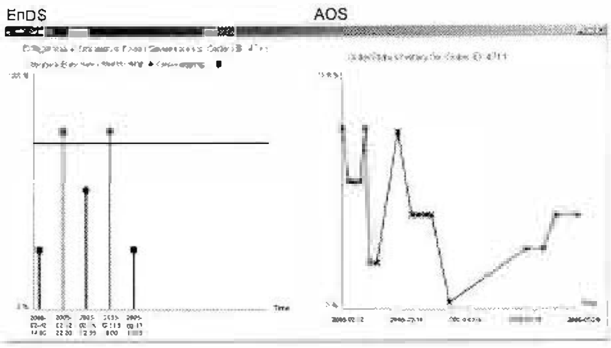


Figure 4. Visualization of EnDS and AOS

This prototype implementation demonstrates the integration of Fuzzy Logic into software agents. It permits intuitive configuration of fuzzy sets and rule bases with a spreadsheet based template. The open-source Fuzzy Logic application programming interface (API) FuzzyJ-API (NRC 2004) for Java is used to realize Fuzzy Logic calculations within the generic prototype. It provides a flexible interface to design and configure Fuzzy Logic applications. Configuration is realized by a MS-Excel file: It specifies all fuzzy variables and fuzzy rules of the application. A Java-Excel-API (Khan 2005) is integrated in the prototype that extracts this configuration data from the MS-Excel file and provides it to the Fuzzy Logic system. This feature offers flexibility for users in maintaining and adapting the analytical behaviors of the SCEM agents.

EVALUATION

Tests

Both, analysis of gathered SCEM data as well as decisions on alerts rely on Fuzzy Logic assessments (see above). In Figure 5 results of tests with the Fuzzy Logic module of the coordination agent are depicted. In these tests several test data sets are analyzed by the coordination agent's Fuzzy Logic behavior. Different strategies are reflected by different Fuzzy Logic rule sets.

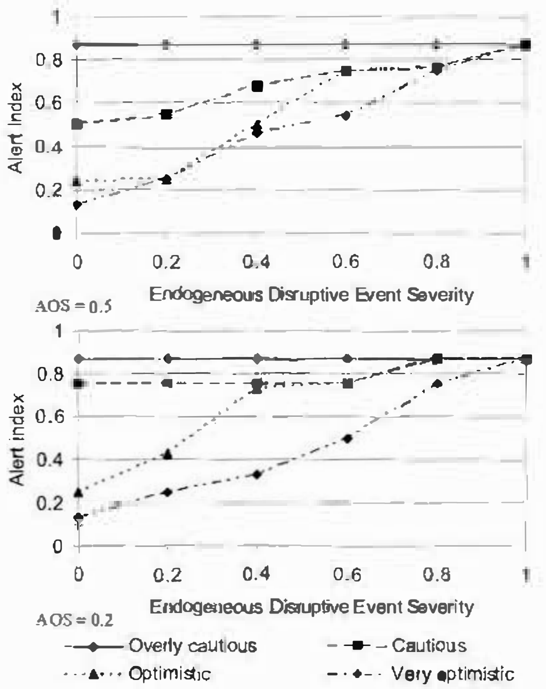


Figure 5. Influence of Fuzzy Logic rule sets

Results of the tests regarding *AOS* and *EnDS* which are both integrated in the *AI* are depicted in Figure 5: *AOS* is fixed and *EnDS* is variable. The same *AI* is calculated with four different Fuzzy Logic rule sets that represent different strategies: Cautious strategies tend to generate alerts even for less severe problems and thus produce higher *AI* than optimistic strategies in the same situation. This behavior is illustrated in Figure 5. For instance, with a medium *AOS* of 0.5 and low *EnDS* an optimistic strategy results in a low *AI*, while a cautious strategy leads to a significantly higher *AI*. This difference increases, if *AOS* is lowered (0.2 in Figure 5, bottom), because cautious strategies value *AOS* higher than *EnDS* and raise *AI* for every disruptive event to a very high level. Optimistic strategies value small disruptive events less, even though the *AOS* is lower.

In addition, experiments with extreme input values are conducted to assess robustness of the system. For instance, a disruptive event with *EnDS=1* (highest possible severity), the lowest possible *AOS=0* and the highest priority of an order *(=1)* is rated with *AI=1*. The results indicate plausible behavior of the Fuzzy Logic components even for these extreme inputs.

Improvement of Configuration

An evaluation of the quality of the heuristic approach is realized with a Fuzzy Logic development tool (XFuzzy 2005). Since no real-world benchmark data is available, realistic assumptions for input and desired output data are provided and tested with the Fuzzy Logic approach.

It is assumed that a human actor can provide consistent heuristic assessments, if confronted with various input data sets. Thus, two input data sets are provided that resemble two similar strategies of two different enterprises for determining an *AI* based on the *AOS* and the *EnDS* (see Figure 6). Specific details of these assessments are of minor importance, only the general structure of the decision graphs is relevant.

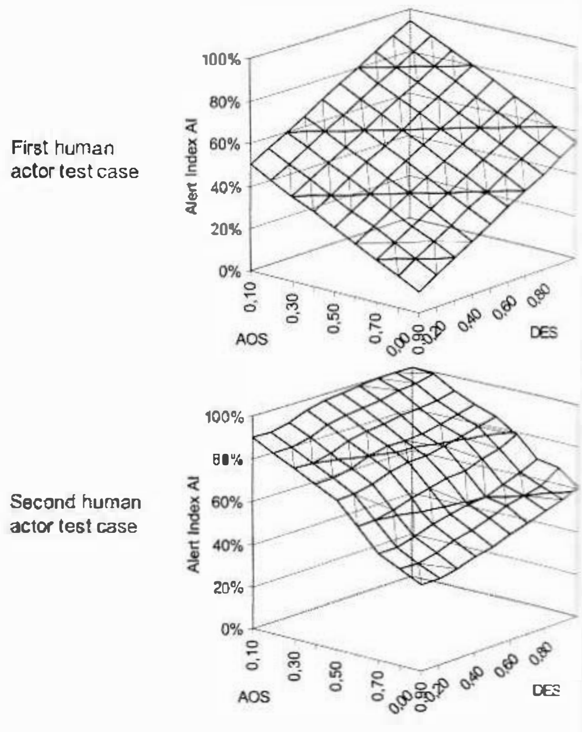


Figure 6. Human assessment samples

Using a simple rule-based system that does not employ Fuzzy Logic would provide a step-like outcome of the *AI*. The result of such a system is also simulated and depicted in Figure 7. The increase of the AI with increasing EnDS and decreasing AOS is realized but not the continuous assessments of the human test cases.

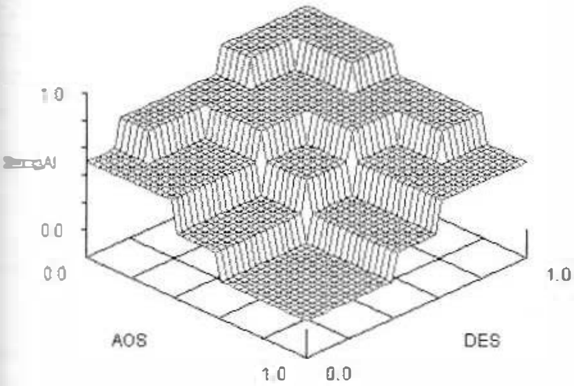


Figure 7. Conventional rule-based system

Figure 7 illustrates binary logic's inability to create continuous assessments with a simple set of rules and functions derived from every day's observations. However, the same rule-base can be used in a Fuzzy Logic system which provides quite different results (see Figure 8). Depending on what types of fuzzy sets are used, a more or less continuous assessment is realized that resembles the first test case while not so much the less symmetric second test case. In several tests "*Pi*" fuzzy sets established the best results which are depicted in Figure 8.

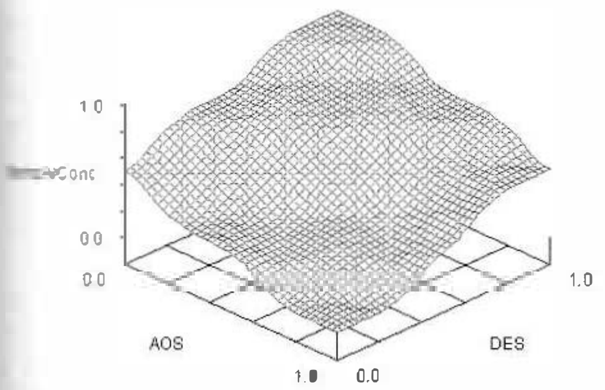


Figure 8. Simple Fuzzy Logic system

To further improve the quality of the Fuzzy Logic assessment and resemble the test cases, the user input is employed as learning material. The fuzzy tool rearranges the fuzzy sets of input und output variables to adapt to the given variable values of the test cases. Several different algorithms are available (i.e. *Steepest Descent, Marquardt-Levenberg, Downhill Simplex, Powell's Algorithm* or *Blind Search*). The aim is to reduce the deviation between test data and Fuzzy Logic assessments to a minimum. The quality is measured by the mean square error (*MSE*). Best results were achieved in experiments with the *Downhill Simplex* (*MSE=0.0015*) and *Powell's* algorithms (*MSE=0.0001*) while a *Blind Search* algorithm, despite its ability to find global maxima, was less successful (*MSE=0.007*). In Figure 9 the result of a successful learning experiment for the second test case is depicted. Compared to the initial human assessment depicted in Figure 6 the similarity of the Fuzzy Logic assessment is very high. Hence, the quality of automated interpretation of SCEM data will resemble a human actor's performance, if configured into a software agent. The optimized configuration is easily extracted from the Fuzzy Logic development tool and configured into the agent-based prototype using the MS-Excel spreadsheet configuration files (see above).

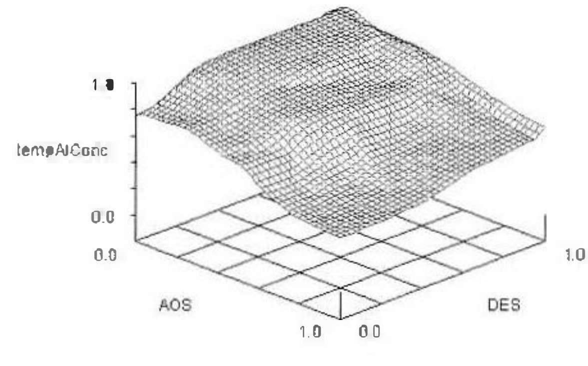


Figure 9. Trained Fuzzy Logic system

CONCLUSIONS

A methodology to heuristically assess data gathered for Supply Chain Event Management purposes is presented and evaluated. Implemented within two types of agents (coordination and surveillance agents) Fuzzy Logic is used to imitate human assessments of complex situations in which an order is situated during its fulfillment. Different strategies a human actor pursues in its interpretations can be adequately reflected by the Fuzzy Logic approach through definition of rule-sets and different fuzzy set types. Besides, automated adjustment of fuzzy sets to learning data is possible. Learning data is derived e.g. from expert interviews.

REFERENCES

Bodendorf, F., Zimmermann, R., "Proactive Supply Chain Event Management with Agent Technology" Int. Journal of Electronic Commerce, 9 (2005) 4, 57-90.

Friedrich, Alfred: Logik und Fuzzy-Logik. 1st edition, expert-verlag, Renningen-Malmsheim 1997.

Khan, Andy: Java Excel API, http://www.andykhan.com/jexcelapi/, accessed on 2005-03-31.

NRC - National Research Council Canada: NRC FuzzyJ Toolkit for the Java Platform, http://www.iit.nrc.ca/IR_public/fuzzy/fuzzyJDocs/, accessed on 2005-04-20.

XFuzzy: http://www.imse.cnm.es/Xfuzzy

Zadeh, Lotfi A.: Some Reflections on the Relationship Between AI and Fuzzy Logic (FL) - A Heretical View. In: Ralescu, Anca L.; Shanahan, James G. (eds.): Fuzzy Logic in Artificial Intelligence. Springer, Berlin 1999, pp. 1-8.

Zimmermann, R., Winkler, S., Bodendorf, F., "Agent-based Supply Chain Event Management – Concept and Assessment", submitted to HICSS 2006.

COMPARING SYSTEM DYNAMICS AND AGENT-BASED SIMULATION TO SUPPORT STRATEGIC DECISION-MAKING

Christian Bauer
Leon Barbulovic-Nad
Freimut Bodendorf
Department of Information Systems
University of Erlangen-Nuremberg
Lange Gasse 20, 90403 Nuremberg, Germany
{Christian.Bauer|Freimut.Bodendorf}@wiso.uni-erlangen.de

KEYWORDS

Approach Comparison, System Dynamics, Agent-Based Simulation, Strategic Decision Support

ABSTRACT

System dynamics and agent-based simulation are used to explore the dynamic behaviour of complex technical and socio-economic systems. Despite different paradigms regarding system representation, model creation and simulation, both approaches have been applied successfully to support strategic decision-making. As the selection of the right modeling approach is elementary for the effectiveness of the decision support, it is necessary to understand approach assumptions and limitations.

This paper contributes to the selection of the right modeling approach by depicting differences between system dynamics and agent-based simulation. First, both approaches and the underlying paradigms are analyzed. Second, it is shown how system dynamics and agent-based simulation offer two different modeling perspectives that carry a different burden of accuracy and model complexity. Explanations are fortified with simplified scenarios and study models describing workforce and knowledge dynamics within an organization.

PROBLEM

In simulation experiments, as well as in real settings, decision makers have access to feedback information about the appropriateness of actions. The closer cause and effect are related, the more effective is the use of feedback information. Unfortunately real decision environments mostly lack this closeness between decision and feedback. It often takes a considerable time until the results caused by a decision are perceptible.

Simulation models compress time and space and thereby enable managers to learn about the effects of decisions more quickly. Using simulation models, decision makers can experiment with various strategies and learn from making rounds of decisions in a environment that allows failure and reflection (Bakken et al. 1994). However, modeling and simulation approaches are based on different methodologies so it is important that model developers and decision makers understand approach differences. They must be aware of resulting model limitations, whether simulation is used as a learning tool, training tool, or as a decision aid.

SYSTEM DYNAMICS

System dynamics applies differential equations to model the system of interest. The approach aims at explaining the system structure that causes an observed behavior. Complex systems are seen as an interlocking structure of feedback loops (Forrester 1976). The system under investigation is decomposed and the causal relationships between the identified elements are revealed.

System dynamics uses stock variables to represent the system states. A stock accumulates the influences it receives over time. The change of state that affects a stock at any point in time is described by flow variables. Flows represent the consequences resulting from actions in the system. While stocks and flows are the basis of system dynamics, auxiliary variables reflect how flows are determined. Auxiliaries are used to represent policies that manage a certain stock by controlling the corresponding inflows and outflows.

Numerical integration is used to compute the behavior of modeled systems. Based on differential equations, time is viewed as continuous. Irrespective of the used integration method, simulation is governed entirely by the passage of time. Often referred to as "time step simulation" (Coyle 1996), a number of steps along the time axis are taken during a simulation run. However the modeler must be aware that the size of the time step influences the simulation accuracy.

AGENT-BASED SIMULATION

In agent-based simulations, individual entities are modeled to interact so that cumulative actions shape the environment that encapsulates this virtual society. An agent may represent an individual but also collectives such as firms or states, or artificial entities (Gotts et al. 2003). Agent types can vary from simple, reactive units to more complex, cognitive agents (Drogoul et al. 2003).

Agent-based models do not have a common formalism. Most formalisms are logic-based but subject to implementation differences. A typical agent-based model consists of agents, interaction environments and governing rules. Agents are usually represented as objects containing internal states and capabilities. Over time, the internal states change due to agent-agent or agent-environment interactions. Simple agent types have capabilities based on predetermined rules of behaviour so interactions are very limited while more complex cognitive agents contain adaptive methods of interaction in form of learning. Such capabilities are usually implemented using evolutionary and genetic algorithms. As interactions occur at discrete points of time a discrete-event view is adapted and implemented either using an event-scheduling or process-interaction approach.

APPROACH COMPARISON

Origin of Dynamics

In system dynamics, the basic building blocks are stocks. Being part of feedback structures they determine the system behavior. The accumulation process captured by stocks is central to the system dynamics approach. Stocks accumulate past events through inflows and outflows. Hence, actual stock value reflects the totality of all past events. This accumulation causes inertia. Assuming limited flow rates, the stock value determines how fast a given state can be changed. Creating a delay stocks absorb the difference between inflows and outflows. Stocks are often used as buffers decoupling outflow rates against fluctuating inputs so stock values vary. As decision making is based on stock information, varying stocks often lead to disequilibrium dynamics being characterized by erratic system states.

In case of agent-based models, the dynamics is due to agent-agent and agent-environment interactions. Agents are the basic building blocks. Following specified rules, agents interact within their environment and thus generate the overall system behavior due to emergence. The macro-level system behavior is a result of the micro-level interactions of individual heterogeneous agents. Events that trigger reactions are the source of dynamics in agent-based simulation models. Properties of reactions are specified by the interaction rules. As these interactions occur at discrete points of time, scheduling of events is of great importance for the emergence of the overall system behavior.

Top-Down vs. Bottom-Up Perspective

System dynamics modeling follows a top-down systems view. The high-level structure of the system is sketched providing a conceptualization of aggregate key elements and relationships. Usually the main stocks in the system are identified first, followed by the flows and the relationships that determine the flow rates. During model development initial stocks are gradually decomposed until all relevant feedback loops are captured. System dynamics seeks an endogenous explanation for a given phenomenon based on the identification of dominant feedback structures. Models developed capture emergence by modeling the phenomenon itself (Schieritz and Milling 2003).

The level of necessary model detail depends on the purpose of the model. As an example, to capture the workforce dynamics of an organization it is common to split the stock representing the company's overall employees into a promotion chain. Applying the aging chain archetype, a structure used to model items that are age-dependent, the promotion chain represents different levels in the staff hierarchy of an organization (Sterman 2000). Figure 1 depicts the structure of a two level promotion chain.

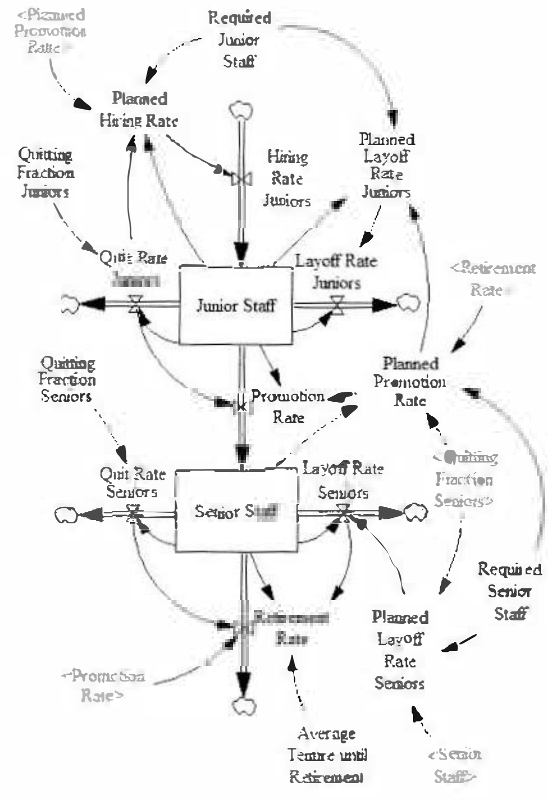


Figure 1: Capturing Workforce Dynamics Using a Promotion Chain Structure

In contrast to the top-down approach, agent-based models follow a bottom-up approach. Individuals are the most basic modeling units. The behavior of individual agents is modeled. Agents usually play different roles and it is possible that one agent is assigned multiple roles or the same agent changes roles during existence. Interactions occur according to the specified interaction rules. Groups of agents can interact with other groups creating a new level of emerging dynamics that is seen as the society behavior. In other words, characteristics of the population evolve during simulation. The bottom-up approach is a source of emergence due to interaction among agents on a particular hierarchical level – the emergence on one level causes an emergent behavior on the level above and so on.

To capture different types of agents, roles and environments, Parunak and Odell suggest the use of an UML class diagram and swimlanes (Parunak and Odell 2002). Relations are shown in a table where vertical swimlanes specify group aggregation while horizontal swimlanes specify object instantiation. An example of such a class-swimlanes diagram is shown in Figure 2.

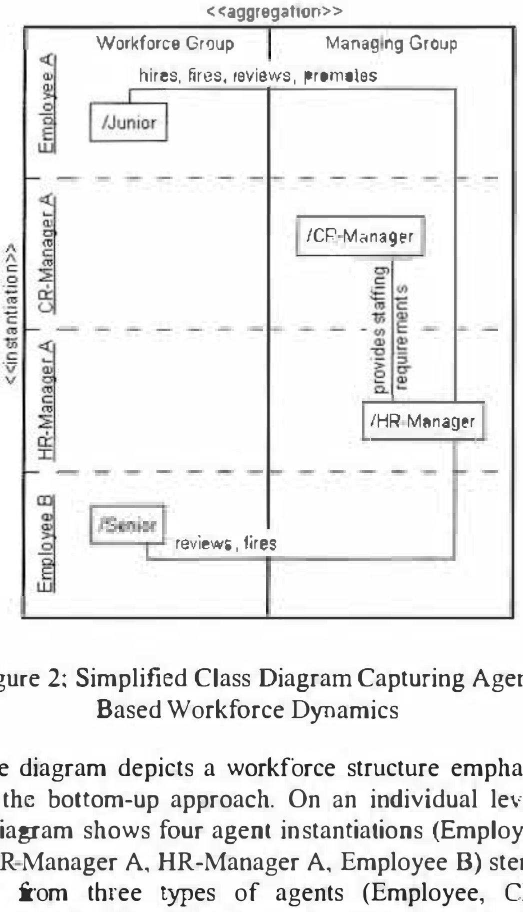


Figure 2: Simplified Class Diagram Capturing Agent-Based Workforce Dynamics

The diagram depicts a workforce structure emphasizing the bottom-up approach. On an individual level, the diagram shows four agent instantiations (Employee A, CR-Manager A, HR-Manager A, Employee B) stemming from three types of agents (Employee, CR-Manager, HR-Manager). Four roles are assigned to the agents (Junior, Senior, CR-Manager, HR-Manager). At the same time, relations between different roles are shown (e.g. the CR-Manager provides staffing requirements to the HR-Manager, an agent who hires, fires, reviews and promotes employees). According to their role agents are organized into two groups: junior and senior employees are members of the Workforce Group while the CR-Manager and the HR-Manager are members of the Managing Group.

Homogeneous vs. Heterogeneous View

Single objects flowing through a stock and flow network cannot be identified and traced in system dynamics. Stocks only represent the quantity of items contained so coflows are used to model the attributes of items in stocks. Coflow structures mirror the main stock and flow network (Sterman 2000). Attributes are modeled using corresponding stocks that keep the attribute values of the items represented in the main structure. Referring to the workforce model (Figure 1) a coflow is used to represent the knowledge of junior and senior staff (Figure 3).

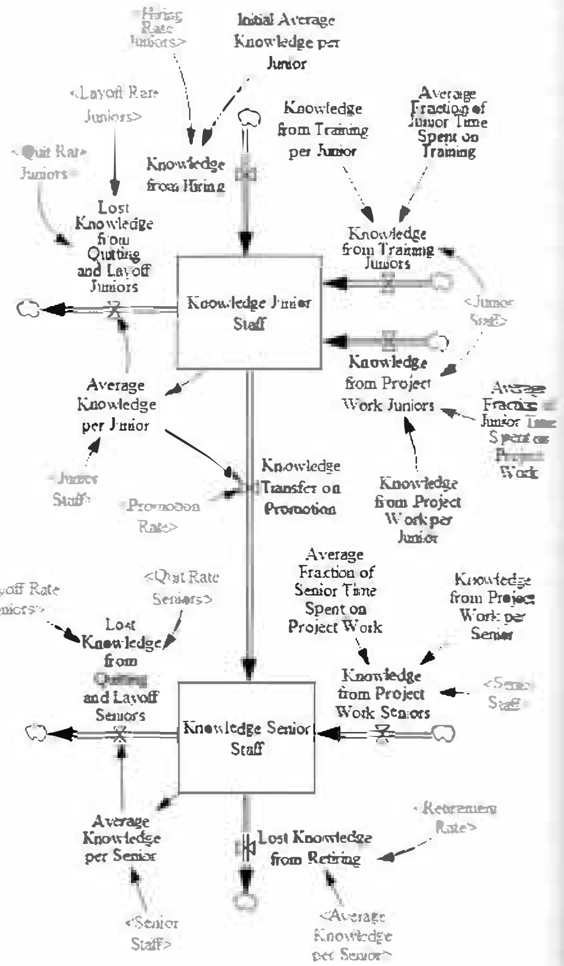


Figure 3: Modeling Workforce Knowledge Using a Coflow Structure

Two stocks are used to represent the knowledge levels of junior and senior members. Further, three inflows are modeled that contribute to the knowledge of junior staff. The first inflow is based on the hiring rate as each new employee brings in a certain amount of initial knowledge. The second inflow is due to project involvement and depends on the average time juniors spend on the project-related work. Further, training is assumed as the third source of knowledge acquisition since juniors not assigned to a project undergo training activities.

The outflow of junior staff knowledge depends on the corresponding layoff, quit and promotion rates. In all cases, the departing juniors take the average knowledge with them. In case of quitting and layoff, the knowledge is lost while in case of promotion, knowledge is transferred to the stock of senior knowledge. It is assumed that seniors are not subject to training activities so knowledge is solely acquired through project work. Similar to the juniors stock, seniors take an average amount of knowledge with them when leaving.

Modeling knowledge dynamics depicts the homogeneous perspective as an inherent aspect of the system dynamics approach. Although the coflow structure captures the overall knowledge on each job level, it is not possible to keep track of individual's knowledge attribute. Hence, the average knowledge per employee is calculated dividing the overall knowledge by the according number of staff members.

On the other hand, agent-based models provide a way of representing heterogeneity. Due to the bottom-up approach agent-based models preserve the individuality observed in real-world systems. Individual agents can have different attributes and rules of behavior depending on their internal states and the surrounding environment. Applied to knowledge modeling, this heterogeneous individual-based perspective provides the opportunity to model individual knowledge sets for each agent. For example, instead of assuming an average knowledge level for every hired junior employee, it is possible to assign individual knowledge sets thereby creating a heterogeneous agent population. Furthermore, the development of individual knowledge sets can depend on agent interaction with other knowledge repositories in the environment. Using unique sets of natural numbers to represent distinct knowledge items, comparison and modification of individual items within knowledge repositories is possible.

Figure 4 depicts various agent interactions with knowledge repositories during project and knowledge-management activities. The figure shows how two agents A and B mutually interact or manipulate project and organizational knowledge repositories by transferring or creating knowledge items. A dashed arrow symbolizes a flow of knowledge from the source to the destination. A solid arrow represents knowledge creation where the tail depicts the origin of the newly created idea while the head identifies the location where the created knowledge is temporally or permanently stored.

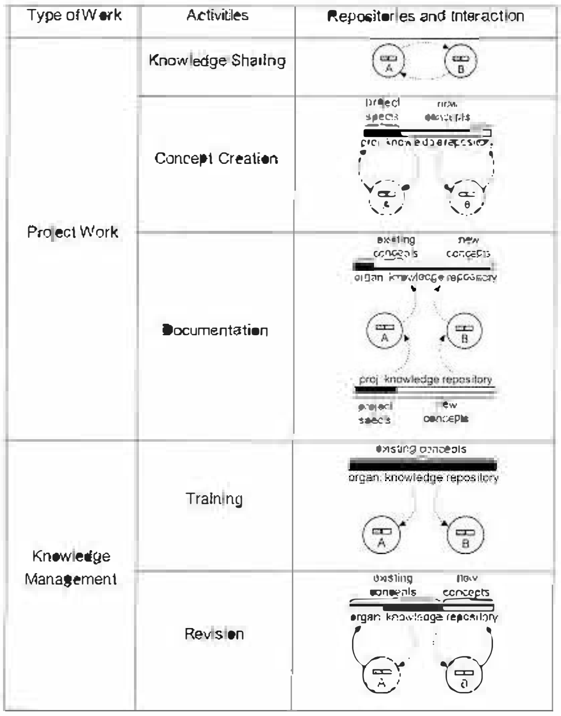


Figure 4: Modeling Knowledge Repositories and Agent Interaction

During project work three activities contribute to the knowledge creation and transfer. First, agents being assigned to a project share some of their individual knowledge while collaborating. Second, the project itself is regarded as a source of knowledge. Agents are confronted with new topics and problems, two additional sources that enlarge their knowledge repositories. While solving problems at hand, agents create new concepts contributing to the project knowledge repository. Third, it is assumed that agents spent some time documenting project findings while transferring the new knowledge items to the organizational knowledge repository.

Agents not involved in project work spent their time on knowledge management activities. Training time is allocated to allow juniors to read and accept approved concepts that are documented as part of organizational knowledge repository. Seniors spend time reviewing already documented concepts while creating new ideas – concepts are merged and extended into new knowledge items. However, all knowledge transfer and creation rates have an upper limit as a way of representing an agent's cognitive limits and the resource use restrictions. The individual's learning capabilities, group contributions to ongoing projects, project contributions to organizational knowledge, as well as knowledge management efforts such as training and revision are limited.

Model Accuracy

To elaborate on model accuracy, the simulation results of the simple workforce and knowledge models, which have been introduced above, are presented. Continuous workforce growth is assumed for the entire simulation time of ten years. In order to ensure equivalent initial conditions, average initial knowledge levels of the instantiated junior and senior agents are used for parameterization of the system dynamics model. The quit rates and the retirement age are set the same. Figure 5 shows the simulation results of both models for the average junior knowledge.

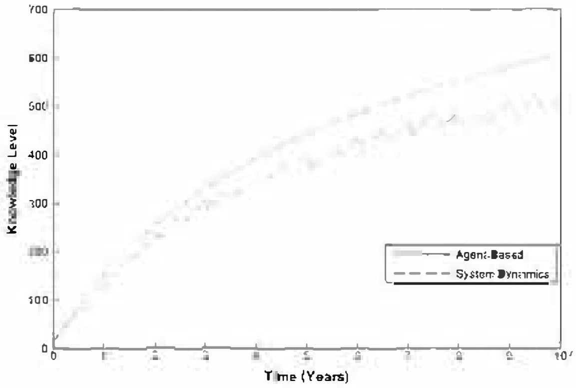


Figure 5: Average Junior Knowledge

The results reveal that compared to the agent-based simulation, the system dynamics model overestimates junior knowledge values. Due to the aggregate and homogeneous modeling perspective of the approach, an average knowledge value is removed from the junior knowledge stock in cases of quitting, layoff and promotion. However, the use of the average knowledge is clearly an oversimplification of the amount of knowledge that is lost or transferred to the senior level.

Although the system dynamics approach cannot handle individual items, the use of an estimated attribute distribution over the total amount of items leads to more sophisticated results. Instead of assuming that juniors possess the same average amount of knowledge, a uniform distribution is implied. Minimum knowledge is calculated using a moving average of the initial average knowledge per hired junior. The time span taken for the moving average equals the average tenure of juniors. Maximum knowledge is given by doubling the difference between the average knowledge of all juniors reduced by the moving average of the initial knowledge per junior. Figure 6 shows the simulation results of the improved model versus the agent-based simulation.

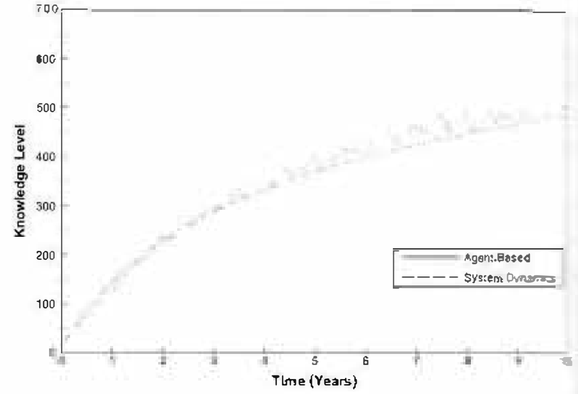


Figure 6: Average Junior Knowledge Using Uniform Attribute Distribution in the System Dynamics Model

Uniform attribute distribution results are in better agreement with the agent-based simulation results. Nevertheless, in a case of heterogeneous attribute values, the agent-based model provides a more accurate result than the system dynamics model. The aggregate and homogeneous view inherent to the system dynamics approach reduces accuracy compared to the individual and heterogeneous view of agent-based simulation.

Model Development

Using simulation models to support managerial decision making, two development aspects are important. First, seriously impinging the applicability of models as decision-making tools, efforts necessary to develop appropriate models are considered complex and time-consuming. Second aspect not to be ignored is the client participation. An effective simulation-based tool requires from the decision makers to accept the simulation model as an adequate representation of the problem. Fostering management participation in model development promotes the acceptance of resulting simulation models (Lane 1994). However, modeling only becomes an active part of decision support if the selected approach is simple enough to allow decision makers to participate in the model formulation and implementation without technical savvy.

Compared to the agent-based models, the development of system dynamics models takes significantly less time. In addition, modeling elements are easy to understand on a qualitative level. Causal loop and stock and flow diagrams allow a graphical model development that reveals the overall model structure. Easy-to-use workbench tools are available for model specification, implementation, execution and documentation as well as visualization and analysis of the simulation results. Nevertheless, modeling and especially the development of mathematical models are tasks left to modeling experts since comprehensive experience is necessary to develop valid models in a reasonable amount of time.

The development of agent-based models proves to be more involved and more time consuming as it is far more concerned with programming details. Standard modeling approaches such as UML, which are increasingly being used for business modeling and modeling of other non-software systems, contributes to the facilitation of agent-based modeling. Advances in graphical representation of agent-based concepts based on UML (e.g. AUML) facilitate the collaboration between developers and end-users. Following well-established standards, models will be easier to use, communicate and understand.

Model Analysis

To gain valuable insights from model analysis, it is important to have an understanding of the model structure and its behavior. Model transparency and the ability to trace the causes of a given behavior are indispensable. The graphical representation of system dynamics models supports this comprehension. The perception and communication of such graphical models is easy and identification of dominant and critical loops is possible. Due to the top-down approach, system dynamics already assures a basic understanding of the model structure as model development starts with an aggregate, hence less complex, view of the system while details are added step by step.

Work with agent-based models is somewhat more difficult. These difficulties are partially due to the bottom-up approach that requires more modeling details but such approach sometimes provides a more intuitive way of system representation. Agents are usually related to corresponding real-world objects or concepts, e. g. persons, machines, orders, etc. While being true regarding the analysis of individual agent interaction, this is questionable for the analysis of emergent behavior on the systems level. Following the bottom-up approach, the number of agents could easily reach into several hundreds, leaving the number of interactions as a multiple. Wooldridge and Jennings conclude that "[...] the dynamics of multi-agent systems are complex, and can be chaotic. It is often difficult to predict and explain the behavior of even a small number of agents; with larger numbers of agents, attempting to predict and explain the behavior of a system is futile" (Woolridge and Jennings 1998, pg. 5).

CONCLUSIONS

System dynamics and agent-based simulation offer two different modeling perspectives that carry a different burden of accuracy and model development time. System dynamics aids the understanding of complex system structures so that effective policies can be designed and targeted toward the most rewarding goals. However, regarding the model accuracy, the agent-based models outperform the system dynamics approach. Based on an aggregate view, system dynamics captures only homogeneous groups of objects whose members are not distinguishable. Agent-based models in contrast maintain individual heterogeneity implementing different agents. System dynamics models run the risk of oversimplification while agent-based models in contrast have to cope with complexity. Complexity reduces the ability to quickly identify relevant behaviors and the corresponding factors of influence. Understanding where the agent-based approach yields additional insight and where details have no importance is crucial in selection of the appropriate method (Rahmandad and Sterman 2004).

REFERENCES

Bakken, B.; Gould, J.; Kim, D.: Experimentation in Learning Organizations: A Flight Simulator Approach. Morecroft, J.; Sterman, J. (Eds.): Modeling for Learning Organizations. Portland: Productivity Press, 1994, pp. 243-266.

Coyle, R. G.: System Dynamics Modelling: A practical approach. London: Chapman & Hall, 1996.

Drogoul, A.; Vanbergue, D.; Meurisse, T.: Multi-Agent Based Simulation: Where are the Agents? Sichman, J. S.; Bousquet, F.; Davidsson, P. (Eds): Multi-Agent-Based Simulation II: Third International Workshop (MABS 2002). Lecture Notes in Computer Science 2581 (2003), pp. 1-15.

Forrester, J. W.: Urban Dynamics. Cambridge: MIT Press, 1976.

Gotts, N. M.; Polhill, J. G.; Law, A. N. R.: Agent-Based Simulation in the Study of Social Dilemmas. Artificial Intelligence Review 19, No. 1 (2003), pp. 3-92.

Lane, D. C.: Modeling as Learning: A Consultancy Methodology for Enhancing Learning in Management Teams. Morecroft, J.; Sterman, J. (Eds.): Modeling for Learning Organizations. Portland: Productivity Press, 1994, pp. 85-177.

Parunak, H. V. D.; Odell, J.: Representing Social Structures in UML. Wooldridge, M.; Ciancarini, P.; Weiss, G. (Eds): Agent-Oriented Software Engineering Workshop II. Berlin: Springer, 2002, pp. 1-16.

Rahmandad, H.; Sterman, J.: Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models. MIT Sloan Working Paper 4512-04, 2004.

Schieritz, N.; Milling, P. P.: Modeling the Forest or Modeling the Trees – A Comparison of System Dynamics and Agent-Based Simulation. Proceedings of the 21th International Conference of the System Dynamics Society (2003).

Sterman, J. D.: Business Dynamics. Systems Thinking and Modeling for a Complex World. McGraw-Hill/Irwin, 2000.

Wooldridge, M.; Jennings, N. R.: Pitfalls of Agent-Oriented Development. Proceedings of the Second International Conference on Autonomous Agents (Agents 98) (1998), pp. 385-391.

DEVELOPING 2D-3D AGENT BASED SIMULATIONS IN GEOGRAPHIC ENVIRONMENT: AN APPROACH AND ITS APPLICATION TO SIMULATE SHOPPING BEHAVIOR IN A MALL

Walid Ali
Bernard Moulin
Computer Science Department
Research Center on Geomatics
3904 Pav. Pouliot, Laval University, Ste-foy, QC, Canada
{Walid.ali, Bernard.moulin}@ift.ulaval.ca

KEYWORDS.
Simulation method, Agent-Based Simulation, Geosimulation, Shopping behavior, Shopping Mall.

ABSTRACT.
Agent Based Simulation (ABS) differs from other kinds of computer-based simulations in that (some of) the simulated entities are modeled and implemented in terms of agents. The agents' capabilities make ABS an attractive simulation concept. Several researchers proposed methods and approaches to develop ABS applications. Unfortunately, none of these methods or approaches takes into account the spatial and geographic aspects of the system to be simulated or those of the simulation environment. In our research we are interested in Agent-Based geosimulation and we focus on the use of the ABS paradigm to build simulations of human behaviors in geographic environments. This paper aims to present a new generic approach that can be used to develop agent-based geosimulation applications which simulate various kinds of system or agents' behaviors in georeferenced virtual environments. In order to illustrate this approach, we present a practical example of an agent-based geosimulation application which simulates the shopping behaviors of customers in a virtual georeferenced world representing a shopping mall. We also show how mall managers can use this type of simulation to make informed decisions about mall's configuration with the objective of making the mall more comfortable to shoppers.

INTRODUCTION

"Agent-Based Simulation (*ABS*) differs from other kinds of computer-based simulation in that (some of) the simulated entities are modeled and implemented in terms of agents (Davidsson, 2000). The agents' capabilities make ABS more attractive than traditional simulation approaches such as Discrete Event Simulation (*DES*), the Continuous Event Simulation (*CES*), and the Object Oriented Simulation (*OOS*). Several applications are created using the ABS paradigm. Our research concentrates on the use of ABS to build simulations of human behavior in virtual spatial environments. "The simulation of human behavior in space is an extremely interesting and powerful research method to advance our understanding of human spatial cognition and the interaction of human beings with the environment" (Frank et al. 2001). Several researchers used this paradigm to develop applications that simulate different kinds of behaviors in spatial environments. For example, (Raubal, 2001) and (Frank et al. 2001) presented an application which simulates wayfinding behaviors in an airport. (Dijkstra et al. 2001) used cellular automata to simulate pedestrian movements in a shopping mall. (Koch 2001) simulated people movements in a large scale environment representing a town. In these applications, the spatial features of the simulation environment (SE) are represented using maps, cellular automata, etc. Other researchers, like (Moulin et al. 2003), (Mandl 2000) and (Koch 2001) emphasized the importance of using Geographic Information Systems (*GIS*) to represent the spatial and geographic features of the simulation environment. With a good representation of the spatial features of the SE it is possible to create more plausible simulations of agents' behaviors in spatial or geographic environment such as pedestrian movement, migrations, road traffic, etc. With the multitude of applications that emphasize the spatial features of the SE, some simulation sub-fields appeared such as *spatial simulation*, *urban simulation*, etc. (Benenson and Torrens 2004). Recently, a new form of simulation called *geosimulation* became popular in geography and social sciences in recent years. It is a useful tool to integrate the spatial dimension in models of interactions of different types (economics, political, social, etc.) (Mandl 2000). This form is supported by advances both in geographical sciences and in fields outside geography (Benenson and Torrens 2003). (Mandl 2000), (Koch 2001) and (Moulin et al. 2003) presented MultiAgent-Based Geo-Simulation as a coupling of two technologies: The ABS technology and the GIS one.

- Based on the ABS technology the simulated entities are represented by software agents which autonomously carry out their activities. They can interact and communicate with other agents. They may be active, reactive, mobile, social or cognitive (Koch 2001).

- Using the GIS technology, spatial features of geographic data can be introduced in the simulation. The GIS plays an important role in the development of geosimulation models. New methodologies for manipulating and interpreting spatial data developed by geographic information science and implemented in GIS have created added-value for these data (Benenson and Torrens 2003).

There are several research works dealing with simulation development methods and applications. For example, we can cite (Zeigler 1979) (Fishwick 1991) (Fishwick 1995) (Tuncer et al. 1984) (Curwood and Balderston 1963) (Banks 1998) (Averill and Kelton 2000) (Anu 1997) (Groumpos and Merkuryev 2002) who proposed various simulation methods and applications. As mentioned in (Drogoul et al., 2002), these methods and approaches, although useful when it comes to understand how to design a simulation, have some major drawbacks: (1) they do not specifically address MultiAgent Based Simulation, but rather computer simulation in general; (2) they are mainly task-oriented, rather than model-oriented, and make it difficult to understand the difficulties found in translating conceptual to computational models. (Ramanath and Gilbert, 2003), (Drogoul et al., 2002) presented generic methods and applications for ABS but, unfortunately, none of these methods or applications takes into account the spatial and geographic aspects of the system to be simulated or those of the simulation environment. In our work we are interested in agent-based geosimulation which is a recent concept in the computer simulation fields and especially in ABS. Based on our literature review we did not found any method to support the development of agent-based geo-simulations involving a large number of autonomous agents evolving in a georeferenced or geographic virtual environment (Ali and Moulin 2005).

In this paper we propose a generic approach to develop agent-based geosimulations that simulate various kinds of *spatial* agent behaviors in virtual geographic environments. We mean by *spatial behavior*, a behavior which is performed in a geographie environment and relates to the spatial features of such an environment. In this paper, we illustrate our approach using an application simulating customers' shopping behavior in a mall called *Square One* in Toronto. To simulate the shopping behaviors, which take place in a geographic environment, we need to use a multiagent geosimulation approach in order to take into account the spatial characteristics of hundreds of shoppers' activities (perception, displacements, etc.) in the virtual geographic environment (Ali and Moulin 2005).

The paper is organized as follows. In Section 2 we present the main steps of our generic approach for the development of agent-based geosimulations. In Sections 3 and 4 and in order to illustrate our approach, we show how it enabled us to simulate the shopping behavior of customers in a virtual shopping mall. In Section 4 we present how shopping mall managers can use our shopping behavior simulation prototype to make informed decisions about their mall configuration in order to make it more comfortable and attractive to customers. In Section 5, we present related works and we conclude the paper.

CREATION OF A 2D-3D MULTIAGENT GEOSIMULATION

Figure 1 depicts the main steps of our approach. These steps are illustrated using the shopping behavior simulation application.

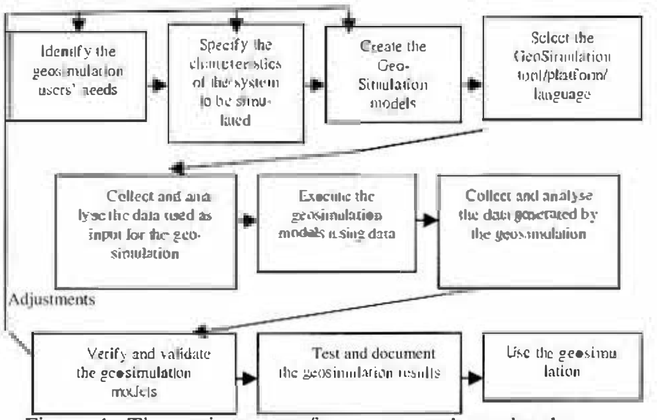


Figure 1. The main steps of our approach to develop multiagent geosimulation

The following sub-sections present the first steps that we propose to follow in order to systematically create geosimulations. The text presenting each step is printed in italics.

Identify the Geo-Simulation users' needs

Simulation applications are generally used to support decision making. In geo-simulation applications decisions are influenced by the spatial characteristics of the simulated system and the geographic features of its environment. Before developing a geo-simulation application we must study in detail the needs and goals of its future users. This step is very important because it helps designers to identify future users, the goals and limits of the system and the goals and limits of the simulation.

In the case of our shopping behavior geosimulation application the users (mall managers) wanted to use the application to simulate and visualize in 2D and 3D customers' shopping behavior in a shopping mall and to assess the influence of different shop locations on the customers' behavior. This evaluation of different mall configurations would help them to make changes in the mall in order to provide customers with a more attractive and comfortable shopping environment. Based on these needs we can limit the context of the geosimulation application to the spatialized *shopping behavior* of customers in a *shopping mall.*

Specify the characteristics of the system to be simulated

Based on the users' needs we must identify the characteristics of the system to be simulated as well as its environment, including all the relevant spatial and non-spatial features within the limits that were defined in the previous step. This step is important because it prepares the ground for the following steps.

In our shopping behavior simulation case we studied the shopping behavior (the system) of people in a mall (the environment). After several months of study of the literature from several disciplines (consumer behaviour, marketing, social psychology, etc.) we got the following results.

The shopping behavior is influenced by several factors:

- *Internal factors*: Demographic (gender, sex, marital status, life cycle, sector of employment, etc.), personality, values, culture, attitudes, habits, preferences, emotional factors. (Duhaime et al. 1996).
- *External factors*: Family, reference groups, social class, etc. (Duhaime et al. 1996).
- *Situational and contextual factors*: The environment ambiance (music, odors, temperatures, etc.) (Deborah et al. 1991), the spatial and geographic configuration of the environment (layout of the stores, textures, color, etc.), and the social aspect of the environment (the attendance of other people, staff, etc.) (Eroglu and Gilbert 1986).
- *Other factors*: The temporal factor (period of time in the day, in the week, in the month, in the year, etc.), expected duration of shopping.

The shopping behavior can be thought of as composed of several processes such as (Petrof et al.,1978): 1) *recognizing shopping motivations*, 2) *information retrieval* used to search for stores where to shop (internal search from the memory or memorization process; and external search in the environment or perception process), 3) *evaluating of alternatives* (choose a particular store), 4) *decision making* before visiting a shop, 5) *post-decision process* (evaluation of the experience after visiting a shop).

The presentation of the details of these factors and processes is not in the scope of this paper. For more details refer to (Duhaime et al. 1996), (Petrof et al. 1978) and (Deborah et al. 1991).

Create the geosimulation models

In order to be able to simulate the studied system using a computer, we must model it as well as its environment, taking into account their spatial and non-spatial aspects. Since our simulation approach is based on the agent technology, we use an agent-oriented design method to create the models and represent the entities of the simulation. The Agent-Based Unified Modeling Language (AUML) (http://www.auml.org/) is an example of such a method. In a simulation we can distinguish two categories of entities: passive and active agents.

The Passive Agent model (PA) is used to specify entities which do not have behaviors. Usually, the most important elements of the simulation environment belong to this category. We must characterize the spatial and non-spatial *structures* of the passive agents. In the shopping behavior simulation case the majority of the shopping mall entities such as stores, kiosks, toilets, doors, entertainment areas, rest areas, smoking areas and parking lots are represented by this category of agents.

- *Non-spatial structure (The Store PA)*: The non-spatial structure of a Store PA contains the information which is specific of a particular store in the virtual SE. For example this structure contains the Store_Identification, the Store_Name, the Store_Speciality, etc. The details of the non-spatial structure of the others PA are not given in this paper.
- *Spatial structure (The PA of the spatial environment)*: The 2D spatial (geographic) structure of the spatial environment PA is modeled using the GIS software GeoMedia (http://www.intergraph.com/). Figure 2.a presents the 2D spatial structure (GIS) of the first floor of the Square One Mall. To create the 3D spatial structure of the PA, we use the software 3DStudioMax (http://www.techanim.com/). A portion of the 3D spatial structure of the first floor of the Square One Shopping Mall is displayed in Figure 2.b. To make our simulation environment more realistic we used pictures of stores windows as textures that were put on the facades of the stores in the virtual

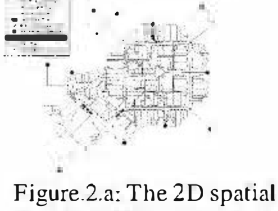


Figure.2.a: The 2D spatial structure of the simulation environment agents

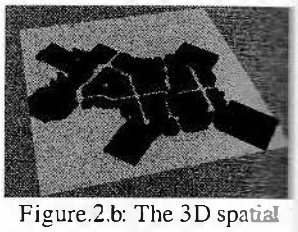


Figure.2.b: The 3D spatial structure of the simulation environment agents

The Active Agent model (AA) is used to specify entities having behaviors. These entities actively participate in the simulation. In this model we specify the *data structures* of the entities (spatial and non-spatial structures) and their *behaviors* (spatial and non-spatial behaviors). In the shopping behavior case we only consider one category of agents which represent the shoppers. *The non-spatial structure (The Shopper AA)*: In the non-spatial structure of the Shopper agent we take into account the shopper's characteristics which can influence the shopping behavior in a shopping mall. We specify the agent's demograpbic profile (identification, name, gender, age group, etc.), preferences, habits, shopping goals, emotional states, as well as dynamic variables (hunger, thirst, etc.), possession state (what the agent owns), agent's knowledge (what the agent knows in the mall: the stores, the toilets, etc.), etc. A dynamic variable is a variable whose values change during the course of the simulation. Dynamic variables can be used to activate certain behaviors. For example, when the need to go to the toilet reaches a certain threshold, the goal of finding a toilet becomes a priority (Moulin et al. 2003))

- *The spatial structure (The Shopper AA)*: The spatial structure of the Shopper AA depicts the spatial representation of the agent in the simulation environment. For example, in the 2D simulation, the spatial structure of the Shopper AA can be a point, a circle or a square. In the 3D simulation, we represent the agents' spatial structure using a 3D shape (a 3D mesh) which represents a young man/woman, an old man/woman; we can choose the colors of clothes.

- *The non-spatial behavior*: In the non-spatial behavior of the Shopper AA are included the main processes of the shopping behavior which are not related to the external environment such as the needs detection process, the internal information retrieval process. *The spatial behavior*: The spatial behavior of the shopper agent depicts the agent's interactions with the simulation environment (movement, obstacle avoidance, path finding, etc.). For example, in a 2D spatial behavior we can see the agent move from one location to another. In a 3D spatial behavior, and using a 3D mesh animation, the agent "walks" in the 3D model of the shopping mall.

Select a geosimulation tool/platform/language

During this step the designer must choose the simulation tool, platform or language that will be used to execute the simulation models. Several simulation tools, platforms and languages can be used to simulate systems or behaviors. Naturally, a question arises: How to select the appropriate simulation platform or tool for a given application? Metrics to evaluate simulation tools include modeling flexibility, ease of use, modeling structure (objects, agents, etc.), code reusability, graphic user interface, animation, hardware and software requirements, output reports, customer support, and documentation.

For our shopping behavior simulation we used a generic platform called MAGS (Multi-Agent Geo-Simulation) (Moulin et al. 2003) which can be used to develop agent-based geosimulations that involve a large number (thousands) of autonomous software agents interacting in virtual geographic environments (VGE). In MAGS the agents are capable to perform "cognitive" and spatial activities in the VGE because they perceive the objects contained in the virtual spatial environment, the features of the ground, paths, roads, buildings and other static objects as well as the other moving agents (Moulin et al. 2003). They are also able to navigate autonomously in the simulation environment based on their perception and memorization mechanism (Perron and Moulin 2004). They also make decisions based on their goals and sub-goals whose priority varies in function of their needs (physiological, social, emotional, etc.) represented by dynamic variables. The needs may be prioritized according to Maslow's classification (Maslow, 1970).

Collect and analyse the data used as input to the Geo-Simulation

In this step we collect data and transform it in order to feed the simulation models. If it is acceptable to input random data in the simulation models, this step can be very simple but the simulation may be unrealistic. But, if we want to use real data we must collect and analyse it before feeding it in the system. Since we deal with geosimulations, we must collect and analyse both non-spatial and spatial data. In our approach we use OLAP and SOLAP techniques to analyse the input data.

For the shopping behavior simulation case and in order to have a realistic simulation we used real data that our team collected in the Square One shopping mall. In this section we briefly explain how we collected the data and which techniques we used to analyse it.

The data collection: The data characterizing the spatial environment is recorded in a GIS and obtained after processing different documents: maps, descriptions of stores, etc. For the creation of the shopper agents we did not have any data. Consequently, we decided to build a survey to collect data about real shoppers visiting the shopping mall. Thanks to this survey that was conducted in the Square One mall during October 2003, we collected about 390 filled questionnaires. In these thirty-pages questionnaires we collected a lot of non-spatial data (customer's demographic profile, habits, interests and preferences) and spatial data about the shopper spatial knowledge (preferred entrance doors, preferred parking lots, usual paths followed during the shopping trip, the shopping areas which are best known in the shopping mall). The data has been collected on paper questionnaires. In order to digitalize this data, and using Microsoft Visual Basic, we programmed a software that is used to input non-spatial and spatial data about the shoppers into a Microsoft Access database.

The data analysis: OLAP and SOLAP analysis: The survey provided lots of non-spatial and spatial data which must be analysed. To do this analysis we used a multidimensional analysis approach based on *On Line Analysis Processing* (OLAP) for the non-spatial data and on *Spatial On Line Analysis Processing* (SOLAP) to analyse the collected spatial data (Bédard and al. 2001). OLAP-SOLAP approach is geared towards decision-support as it is designed from the start to be *easy* and *rapid* (Rivest et al. 2001).

For the shopping behavior simulation case, an example for the OLAP and SOLAP analysis results is presented in the two following points.

- *OLAP analysis*: Using OLAP analysis we can analyse non-spatial variables which are called *Dimensions*. We can also determine the influence of one dimension on another. For example, we can determine the influence of the gender dimension on the color or music preferences dimensions. Actually, we analysed results about all the dimensions of our model of the Shopper agent. We can further analyse the data by combining dimensions together.
- *SOLAP analysis*: Using a SOLAP analysis we can determine the relationship between a spatial dimension of the environment and the non-spatial dimension of the Shopper agent. For example, we can determine the relationship between the Gender dimension of a shopper and the choice of the shopping corridor or the shopping floor in a shopping mall. Figure 3.a presents the entrance doors of the first floor of the Square One shopping mall and Figure 3.b presents the graphical representation of the distribution of the participants on the dimension Floor_Entrance_Door in the shopping mall. We can see in Figure 3.b that the most frequented mall's doors are Door 0 (97 shoppers) and Door 10 (125 shoppers).

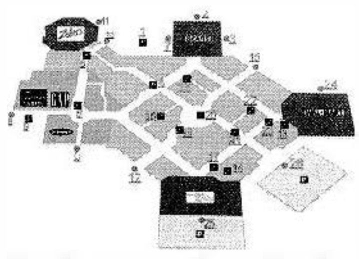


Figure. 3.a: The entrance doors of the first floor (Square One Mall)

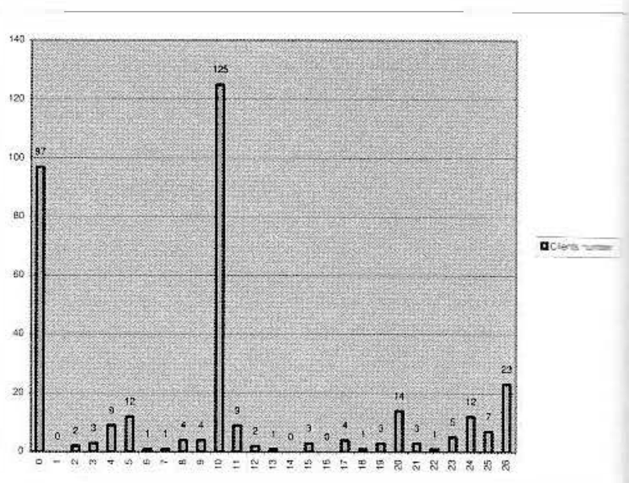


Figure. 3.b: The distribution on the Floor_Entrance_Door dimension

RUN & VALIDATE THE GEOSIMULATION

In this section we present the steps of our method which consist in running the geosimulation, collecting relevant data generated by the simulation, and validating the geosimulation models, and using the geosimulation for decision making.

Execute the geosimulation models

During this step we implement the simulation models on the selected simulation tool/platform/language using the data characterizing the system and its environment. During this step we must respect the constraints and limits of the selected tool/platform/language such as the input data structures.

To develop our shopping behavior simulation, we used the MAGS platform. For each simulation we must prepare a simulation scenario. In such a scenario, we must indicate for example which percentage of shopper agents (with specific characteristics) enter at each door at given times. To create this scenario we use a dedicated interface that belongs to the MAGS platform.

After the simulation scenario preparation we can execute the simulation models in the MAGS simulation engine. Figures 4.a and 4.b present respectively 2D and 3D screen of the simulation execution using MAGS. In this figure we can see the shopper agents who navigate in the environment and visit stores to achieve their shopping goals.

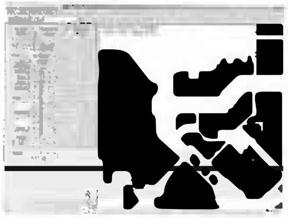


Figure.4.a: A 2D simulation execution: The shopping behavior in the Square One shopping mall (Toronto)

Figure.4.b: A 3D simulation execution: The shopping behavior in the Square One shopping mall (Toronto)

In the simulation prototype the Shopper agent comes to the mall to visit a list of specific stores or kiosks (by name or type) that are chosen on the basis of the agent's characteristics and of the data collected during the survey. It enters by a particular door and starts its shopping trip. Based on its position in the mall, on its spatial knowledge (memorization process) and on what it perceives in the mall (perception process), it makes decision about the next store or kiosk to visit (decision making process). When it chooses a store or kiosk, it moves in its direction (navigation process). Sometimes, when it is moving toward the chosen store or kiosk, the agent can perceive another store or kiosk (perception process) that is in its shopping list and that whose location was not in its memory. In this case, the Shopper agent moves to this store or kiosk and memorizes it (memorization process) for its next shopping trips. The shopper agent accomplishes this behavior continually until it visits all the stores or kiosks which were on its visit list or until it runs out of time for its shopping trip. If the shopper agent has still time for shopping and some stores or kiosks of its list are in locations unknown by the agent, it starts to explore the shopping mall to search for stores or kiosks (exploration mode). When the shopper agent reaches the maximum time allowed to the shopping trip, it leaves the mall.

The Shopper agent can also come to the mall without a specific list of stores or kiosks in order to explore it. In this exploration mode the Shopper agent follows its preferred paths in the shopping mall. In this mode the moving action of the Shopper agent to the stores, kiosks, music zones, odor zones, lighting zones, is directed by its habits and preferences. For example, if the Shopper agent likes *cars* and it passes in front of a car exhibition, it can move and stop to attend this exhibition. To extend our simulation prototype we can simulate the shopper reactions to the mall's atmosphere. We can insert special agents that broadcast music, lighting or odor in specific areas. When the shopper agent is in the exploration mode; if it perceives and likes the music or the lighting or the odor broadcasted in these areas, it can move toward them and possibly enter the store increasing its level of satisfaction.

During its shopping trip the Shopper agent can feel the need to eat or to go to the restroom (simulated by a dynamic variable reaching a given threshold). Since these needs have a bigger priority than the need to shop or to play, the agent temporarily suspends its shopping trip and goes to the locations where it can eat something or to the restrooms. In our geosimulation prototype the priorities of the activities of the shopping behavior are defined on the basis of Maslow's hierarchy of needs (Maslow 1970).

Collect and analyse the data generated by the Geo-Simulation

To be useful, the simulation application must return meaningful results. Based on these results and on the analysis of these results, the users can make decisions. In our approach and in order to analyse the simulation output data, we use the same OLAP and SOLAP analysis techniques that are used to analyse the simulation input data. It is important to indicate that the simulation output data is generated by a specific type of software agents: Observer agents.

In the example of the shopping behavior simulation we collect data from the models execution thanks to special software agents called *Observer Agents*. The role of an observer agent is to collect data from the shopper agents which come nearby (in the observer agent's perception range) and to store this data in files or databases. When the simulation execution ends, we can analyse the contents of these files or databases in order to make a report about the simulation results. We can collect non-spatial and spatial results during the simulation. Using again our OLAP and SOLAP techniques and tools we can analyse these results. For the shopping behavior simulation example, we designed the structure of the Observer Agents that collect the simulation results and then we worked on the analysis of these results. For example, Observer agents located at the entrances of the virtual shopping mall record the number of Shopper Agents entering and exiting the shopping mall. Other Observer Agents count the number of Shopper Agents going through specific areas. Other Observer Agents collect different data such as the Shopper Agents' satisfaction when exiting the shopping mall. This Observer agents' activity is similar to conducting a survey in the virtual environment of the same kind as the one we conducted in 2003 in the real shopping mall. Hence, we are able to use the same OLAP and SOLAP analyses that we used to analyze the data of the survey conducted with real shoppers.

Verify and Validate the Geo-simulation models

During this step we can compare the simulationmodels under known conditions with the system that we simulate. This step not only ensures that the model assumptions are correct and consistent, but also enhances the users' confidence in the simulation models (Anu 1997).

Based on the simulation input data and the simulation results we can verify and validate our simulation models.

In the case of shopping behavior simulation, we did not perform this step yet. To this end, we are investigating various verification and validation techniques to verify and validate simulation systems. These techniques have to be customized for geo-simulations. Currently, we plan to use the technique of validation by comparison between the simulation models and the real system which is simulated. We plan to collect data about real shoppers in another mall and compare them with the Agent-Based Geo-Simulation output results in order to validate the shopping behavior models.

Test and document the Geo-Simulation

During this step we document and test the simulation. In the documentation we present the results of the system analysis, the simulation models, the selected tool/platform/language, a guide to use the simulation interface, the input/output data analysis results, etc.

Currently, we have documented a large part of our shopping behavior Geo-Simulator. We hope to complete this documentation and to deliver it to the final users (the Square One shopping mall managers) in the coming months.

Use the geosimulation for decision making

The last step of our approach is to exploit the results of the multi-agent Geo-Simulations in order to for example:

- Understand the system to be simulated by observing various simulations carried out over long periods of time using the GeoSimulation platform.

- Compress time to observe a system over long periods or expand time to observe it in details. To this end, the user can control the simulation time step.

- Experiment the system in new situations or contexts in order to assess the influence of different decisions.

Our shopping behavior simulation can be used by shopping mall managers to make decisions related to *the spatial configuration of the shopping mall.* The shopping mall manager can change the spatial configuration of the virtual mall (change a store features or position, close a door or a corridor, etc.). For each change he runs the simulation and collects the results. By comparing these results he can make decisions about the best spatial change or configuration for the shopping mall. How to propose a systematic way to carry out these comparisons is still an open research area.

To illustrate the use of the Shopping behavior geo-simulation tool we used 2 simulation scenarios. In the first one we launch a simulation with a population of 390 shoppers similar to the population of customers interviewed during the October 2003 survey (Figure. 5.a). This first scenario generates for us output data about the routes that the Shoppers Agents follow in the shopping mall. In the second scenario we exchange the location of two department stores: *Wal-Mart* and *Zellers* (Figure. 5.b). We launch the simulation again and we generate the output data about the routes of the same population of Shopper Agents. By comparing the output data of the two scenarios we notice the difference of the paths that the Shopper Agents followed to attend *Wal-Mart* and *Zellers* stores (Figures 5.a and 5.b). In these figures the flow of the Shopper Agents which pass through a corridor is represented by a line which is drawn in the middle of this corridor. The width and the color of this line are proportional to the flow of Shoppers agents that pass through the corridor. If this flow grows, the width of the line grows and its color becomes darker. The simulation output analysis shows that corridor X is less frequented in scenario 2 than in scenario 1 (Figure. 5.a). However, corridor Y is more frequented in scenario 2 than in scenario 1 (Figure. 5.b). By a data analysis on the characteristics' dimension of the Shopper Agent we can see that in scenario 2, most of the Shopper Agents that go through corridor Y are female and they come to the mall to visit female cloth stores. If the mall manager chooses the mall configuration of scenario 2, he may think of renting the spaces along corridor Y to female cloth stores.

The data analysis of the geosimulation output (non-spatial and spatial data) is implemented in an analysis tool that we developed using *Microsoft Visual basic 6.0*. This user-friendly tool uses the data generated in "Output files" by the *Observer* Agents. Then, based on a "Dimension file" which contains a hierarchy of dimensions, it computes measures for one dimension (i.e., Gender or Age group) or for a combination of dimensions (i.e. Gender and Age group) using an OLAP (On Line Analytical Processing) analysis approach. These measures are recorded in "Analysis files". Then, the tool's interface uses the data contained in the analysis files in order to display multidimensional non-spatial and spatial results. Non-spatial analysis is the result of crossing non-spatial dimensions related to shopper agents. Spatial analysis is the result of crossing non-spatial dimensions with the position of each Observer Agent (corridor) which represents a spatial dimension.

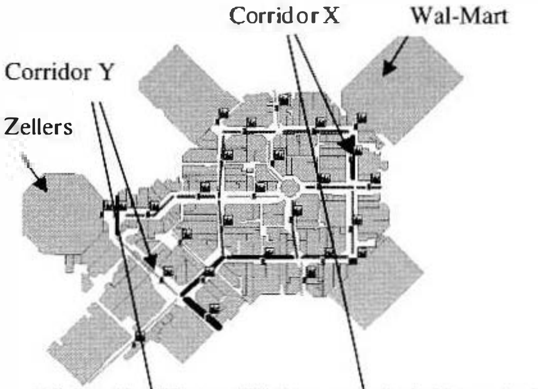


Figure. 5.a: The spatial data analysis in Scenario 1

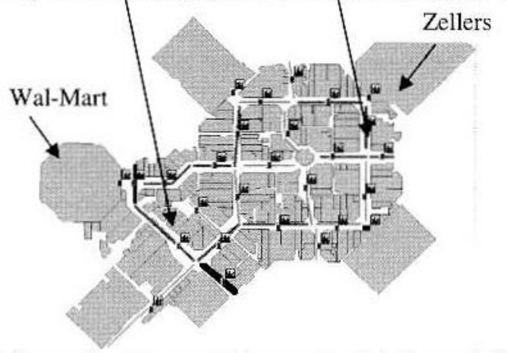


Figure. 5.a: The spatial data analysis in Scenario 2

RELATED WORKS AND CONCLUSION

Some systems such as as traffic systems, urban systems are spatially explicit (objects are associated with locations in geographic space) and exhibit mobility (they move around in the environment). It is relevant to simulate them using the ABS paradigm. Unfortunately, we cannot follow one of the methods or approaches proposed by (Zeigler 1979) (Fishwick 1991) (Banks 1998) (Anu 1997) (Groumpos and Merkuryev 2002) (Ramanath and Gilbert, 2003) (Drogoul et al., 2002) to simulate them. The reason is that these methods do not take into account the spatial and geographic aspects of the system to be simulated or those of the simulation environment. In our work, we are restricted to a recent simulation concept called geosimulation and we propose a generic approach which can be followed to develop 2D-3D Agent-Based applications that simulate some behaviors which are performed in geographic environments. What distinguish our approach is that we take into account the geographic characteristics of the agents and their virtual environment when we develop the simulation. These characteristics are not taken into account (or at best only partially) by existing methods despite their importance in the system to be simulated.

To illustrate this approach we presented an Agent-Based geosimulation application in which we simulate the customers' shopping behavior in a virtual mall. Our simulation prototype is still under development. Consequently, we did not illustrate some of steps of the proposed approach, but this will be done in the near future.

In few months we hope to go through the two final steps (verify/validate our Agent-Based geosimulation models and test/document the geo-simulation) of our approach and to apply them to our simulation prototype in order to deliver a final simulator to its future users.

In conclusion, our research has a very good potential for innovation. To our knowledge it is one of the most elaborated existing simulations of shoppers behavior in a mall because it takes into account both the agent's cognitive abilities (perception, memorization, decision making) and the spatial behaviors of shoppers when moving in a georeferenced environment.

ACKNOWLEDGE

This research is supported by GEOIDE, the Canadian Network of Centers of Excellence in Geomatics, and by the Defense research center (RDDC Valcartier).

REFERENCES

Ali W., Moulin B. 2005. "Towards a generic approach to develop 2D-3D Multiagent Geo-simulation", to appear in the proceedings of the Agent Directed Simulation (ADS'05) (Part of the 2005 Spring Simulation Multi-Conference (SpringSim'05)), San Diego, California (USA). 2-7.

Anu M. 1997. "Introduction to modeling and simulation", Proceedings of the 29th conference on Winter simulation, p.7-13, December 07-10, Atlanta, Georgia.

Averill M. L., Kelton W. D. 2000. "Simulation modeling and analysis". Edited by Averill M. Law, W. David Kelton.

Banks J. 1998. "Handbook of simulation: principles, methodology, advances, applications, and practice". Edited by Jerry Banks.

Bédard Y., Rivest S., Marchand P. 2001. "Toward better support for spatial decision making: Defining the characteristics of Spatial On-Line Processing (SOLAP)". Geomatica, the Journal of the Canadian Institute of Geomatics. Vol. 55, No. 4, p. 539-555.

Benenson I., Torrens P. 2003. "Geographic Automata Systems: A New Paradigm for Integrating GIS and Geographic Simulation". In proceedings of GeoComputation conference, p. 24-30, Southampton, UK.

Curwood A. H. and Balderston F. E. 1963. "Symposium on simulation models: methodology and applications to the behavioral sciences". Edited by Austin Curwood Hoggau and Frederick E. Balderston.

Davidsson P. 2000. "Multi Agent Based Simulation: Beyond social simulation". In Moss, S. and Davidsson, P. (Eds.) *Multi Agent Based Simulation*, Springer Verlag LNCS series, Vol. 1979.

Deborah J., Macinnis C., Whan P. 1991. "The Differential Role of Characteristics of Music on High- and Low-Involvement Consumers' Processing of Ads". The Journal of Consumer Research. Vol. 18, No. 2, p. 161-173.

Drogoul A., Vanbergue D., Meurisse T. 2002. "Multi-Agent Based Simulation: Where are the Agents?". In Proc. of MABS'02 (Multi-Agent Based Simulation, Bologna, Italy, July 2002), LNCS, Springer-Verlag.

Dijsktra J., Harry JP., Bauke U. 2001. "Virtual reality-based simulation of user behavior within the built environment to support the early stages of building design". In M. Schreckenberg and S.D. Sharma (ed.): Pedestrian and Evacuation Dynamics. Springer-Verlag, Berlin. p. 173-181.

Duhaime, Kindra, Laroche, Muller. 1996. "Le comportement du consommateur". 2ème édition. Gaétan Morin, éditeur.

Eroglu A., Gilbert D. 1986. "Retail Crowding: Theoretical and Strategic Implications". Journal of Retailing, 62 (Winter), p. 346-363.

Fishwick P. A. 1991. "Knowledge-based simulation: methodology and application". Fishwick P. A. and Modjeski R. B., editors. Springer Verlag, 1991.

Fishwick P. A. 1995. "Simulation model design and execution: building digital worlds". Edited by Paul A.

Frank A.U., Bittner S., Raubal M. 2001. "Spatial and cognitive simulation with multi-agent systems". in D. Montello (edt.), Spatial information Theory: Foundations for Geographic Information Science. Springer Verlag, LNCS 2205,124-139.

Groumpos P., Merkuryev Y. 2002. "A methodology of Discrete-Event Simulation of Manufacturing Systems: an overview". Studies in Informatics and Control: with emphasis on useful applications of advanced technology. Volume 11 Number 1, p. 103-110.

Koch A. 2001. "Linking Multi-Agent Systems and GIS-Modeling and simulating spatial interactions-". Department of Geography RWTH Aachen. Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symposium Salzburg 2000, Hrsg.: Strobl/Blaschke/Griesebner, Heidelberg, p. 252-262.

Mandl P. 2000. "Geo-Simulation- Experimentieren und Problemlösen mit GIS-Modellen". In Angewandte Geographische Informationsverarbeitung XII, Hrsg.:Strobl/Blaschke/Griesebner, Wichmann Heidelberg, p.345-356.

Maslow A. 1970. "Motivation and Personality". 2nd ed., Harper & Row,

Moulin B., Chaker W., Perron J., Pelletier P., Hogan J. 2003. "MAGS Project: Multi-agent geosimulation and crowd simulation". In the proceedings of the COSIT'03 Conference, Ittingen (Switzerland), Kuhn, Worboys and Timpf (edts.), Spatial Information Theory, Springer Verlag LNCS 2825, p. 151-168.

Perron J., Moulin B. 2004. "Un modèle de mémoire dans un système multi-agent de géo-simulation", Revue d'Intelligence Artificielle, Vol 18/5-8, Hermes.

Petrof V. 1978. "Comportement du consommateur et Marketing". La presse de l'université Laval.

Ramanath A.M., Gilbert N. 2003. "Toward a Methodology for Agent-Based Social simulation Research". Agent-Based Social Simulation Special Interest Group Meeting (Agent-Oriented Software Engineering and Simulation session), p. 3-24, Barcelona, Spain.

Raubal M. 2001. "Agent-Based Simulation of human wayfinding. A perceptual model for unfamiliar building". PhD thesis. Vienna University of Technology, Faculty of Sciences and Informatics.

Rivest S., Bédard Y., Marchand P. 2001. "Toward Better Support for spatial decision making: Defining the characteristics of spatial on-line analytical processing (SOLAP)". Geomatica Vol. 55, N. 4., p.539-555.

Tuncer I. O., Zeigler B. P., Elzas M. S. 1984. "Simulation and model-based methodologies : an integrative view". NATO Advanced Study Institute on Simulation and Model-based Methodologies: an Integrative View (1982 : Ottawa, Ont.).

Ziegler B. P. 1979. "Methodology in systems modelling and simulation". Proceedings of the Symposium on Modelling and Simulation Methodology, held at the Weizmann Institute of Science, Rehovot, Israel, August 13-18.

AUTHORS BIOGRAPHIES

ALI W. obtained a Master degree in Computer Science from Laval University (Quebec). He is now a PhD Student (supervised by Bernard Moulin since 2001). He is mainly interested in using multi-agent systems to simulate behaviors and phenomena in virtual spatial environments. He is working on a project whose goal is to simulate shopping behavior in shopping malls using geosimulation and multi-agent systems.

MOULIN B. obtained an engineering degree from Ecole Centrale de Lyon (1976), a degree in Economics (University Lyon2) and a PhD (applied mathematics-computer science, University Lyon1 (1979)) all in France. He is currently a full professor at Laval University in the Computer Science Department. He is also one of the main researchers of the Research Center in Geomatics at Laval University. He has been working on software engineering and artificial intelligence for more than 20 years. He is supervising research projects mainly supported by the Natural Sciences and Engineering Research Council of Canada, the FQRNT fund of Quebec, the Network of Centers of Excellence in Geomatics GEOIDE and the Defence Research and Development Center at Valcarier. His main interests are: multi-agent systems, agent-based geosimulation, applications of AI techniques in geomatics, modeling and simulation of software agents' conversations.

DENSITY DEPENDENCE AND THE SURVIVAL OF FIRMS A MULTI-AGENT APPROACH

Pedro Campos[1], Pavel Brazdil[2]
LIACC-NIAAD
Faculty of Economics – University of Porto
R. de Ceuta, nº 118, 6º, 4050 - Porto - Portugal
[1] pcampos@fep.up.pt; [2] pbrazdil@liacc.up.pt

KEYWORDS

Multi-Agent Simulation, Density Dependence Model, Genetic Algorithms.

ABSTRACT

We have used a Multi-Agent Based Simulation to simulate a representative set of firms for a specific type of industries and geographical locations and observed the interaction among them, with special attention to their geographical location and mortality. We defined a variant of the density dependence model to set up the dynamics of the firms in the simulation. By comparing some real and simulated results, and adjusting the chosen parameters with the help of a Genetic Algorithm, we obtain estimates for density limits and other intrinsic parameters not observable in reality.

INTRODUCTION

What are the determinants of the survival of firms? What explains their births and deaths? To investigate the survival of organizations some researchers use the density dependence model, where vital rates of birth and death of firms are dependent on the size of the population, the population density. Spatial components (geographical barriers, localized resource environments, etc.) affect also the evolutionary dynamics of organizational populations in a nonlinear way.

Based on this idea and due to its relevance for organizational survival, we focused our research on certain parameters explained in detail later on (eg. density of firms, age and size), and the impact on mortality of organizations using a Multi-Agent based simulation. We used a calibration process, based on Genetic Algorithms to find the combination of these determinants that minimize the distance between real and simulated aggregate behaviour. We compared vital rates (birth and death and growth rates) of simulated data with real data from Portuguese Textile Industry from 1997 to 2003. The region of Ave has been chosen because it is a geographical cluster where textile industries are concentrated in Portugal. To analyse the impact of different variables in the organizational mortality we applied survival analysis models and examined the regression coefficients.

The remaining of the paper is structured as follows: in the next section we introduce the density dependence model and discuss the determinants of the survival of organizations. In the following sections we present some related works and describe the simulation and calibration processes. Conclusions and future work are presented at the end of the paper.

DENSITY DEPENDENCE

The field of Organizational Ecology (OE) establishes a relation between the society outside organizations and the internal life of organizations ("firm" and "organization" are terms used indistinctively in this paper). OE was introduced by Stinchcombe (1965), Carroll and Hannan (1989, 1992), Hannan and Freeman (1984), among others, who received motivation from sociology and ecology. Recent research determined that density (the number of organizations in a population) affects rates of organizational founding and mortality and formal models of density dependence have been developed. In the density dependence model, vital rates of birth and death of firms are dependent on the population density

According to this model, the survival of a firm depends on the number of firms that are located on its neighbourhood. In most situations, this neighbourhood is not strictly geographic but an "industrial neighbourhood", where firms are considered nearby if they share the same market or resources. In OE, the notion of rational behaviour or profit maximization is not accepted as the main driving motivation for firms. Rather, this role is taken by forces of natural selection and organizational inertia (Wissen, 2004). This theory offers a comprehensive way of studying the evolution of populations of firms. We will introduce some variations in the model, as it seems to be weak in two main aspects: first, the size of the organization is not taken into account in theory, whereas clearly large and small firms have quite distinct characteristics. Second, firms differ not only with respect to size and economic activity, but also with respect to geographical location, that is, spatial heterogeneity (Wissen, 2004). The density dependence model is a single industry model, which means that it ignores the relationship with the environment, including the relationship with other industries. For those reasons, we will give special

attention to the geographical location, the size of the firm and the inter-industry relationships in our analysis.
Two basic forces are responsible for the dependency between the population density and founding (birth) and failure (death): legitimation and competition. Legitimation refers to the degree that a new organizational form is known and accepted in society. The founding rate is proportional to the level of legitimation whereas the mortality rate is inversely proportional to the level of legitimation (opposite associations are observed with respect to competition). The join effects of legitimation and competition explain to a large degree the specific S-shaped form of growth rates over time, because as the population grows, legitimation increases and competition is still very restricted, so that the growth rate increases. At a certain level, the maximum level of legitimation is reached, and competition starts to increase fast. Consequently, the growth rate decreases fast to zero or even becomes negative.

RELATED WORKS

Many authors, (as Albino 2003; Zhang 2003) use Multi-Agent Systems (MAS) for modelling clusters of organizations. Epstein and Axtell (1996) explain that the use of MAS is appropriate when interactions among (heterogeneous) agents are complex and nonlinear and when the space is crucial. Considering the impact of density on the mortality of organizations, Carroll and Hannan's (1989) use simulation models to conclude that density at the time of founding may change the mortality rate of adult organizations. Lomi and Larsen (1998) used cellular automata in a Game of Life-like simulation, based on an Ecological Dynamics of local structure in a nearest neighbour model to simulate organizational populations. The main goal of their work was to study the effects of density at the time of founding on firms growth, decline and death. Here we are concerned with an application to a real industrial context, and assign more relevance to some concepts such as size, location and age. We have compared real and simulated results, adjusting some parameters with the help of a Genetic Algorithm and obtained estimates for density limits and other intrinsic parameters not observable in reality. The observation of the simulated aggregate behaviour and its comparison with reality validates the model and helps to identify the main determinants of organizational survival.
This kind of comparison with "real conditions of organizations" within the density dependence model is original in the scope of the works concerning the simulation of organizations.

TECHNIQUES, MODELS AND DATA.

The Multi-Agent System

Our system MASOS – Multi Agent System for Organizational Survival - is composed by a landscape, agents, and a rule. The landscape is represented by a graph that interconnects the places where firms can be located. We focused our environment in four regions of North-West of Portugal where textile industry is concentrated (showed in Fig.2). One of them, Ave, is often named as an Industrial District or cluster and is the centre of our study.

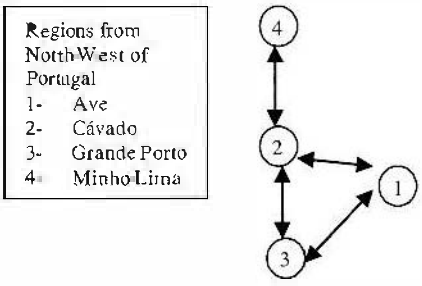


Figure 1: North-western region of Portugal represented as a graph where arrows connect contiguous areas representing the four regions of the study.

Each firm is represented by an agent with a set of attributes: location, age and size. There is one rule in MASOS that connects the agents and the environment which is responsible for the determination of the firms' birth and death processes regarding the variant of the density model presented before.

Variant of the density model

To determine the birth and death of a firm, we have based our representation on a variant of the density dependence model. According to this variant, in every region, a firm can get born or die, depending on a function which is based on the number of firms that are located in its neighbourhood. For that purpose, we have calculated the density, D_j^t, involving a particular region, j, at time t, as the average number of firms per square kilometre that are placed in the region or in the nearest neighbour regions at the period, t:

$$D_j^t = \sum_{i=1}^{K_j} \frac{f_i^t}{Area_i} \qquad (1)$$

In this formula K_j is the number of regions that belong to the neighbourhood of region j; $Area_i$ is the area of the regions involving j, measured in square kilometres and f_i^t corresponds to the number of firms existing in the region i at time t (i=1, 2, ..., K_j), and is computed as follows:

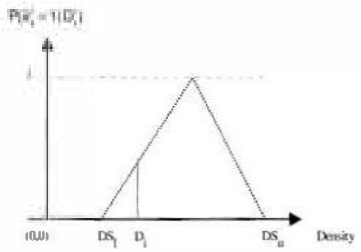
$$f_i^t = \sum_{j=1}^{N_i} \delta_{ij}^t \qquad (2)$$

where N_i represents the total number of firms in the region i and δ_{ij}^t is the state (dead - coded as 0 - or alive – coded as 1) of the j^{th} firm of the region i at time t. For example, if at time 2 the number of firms around a particular firm in the region of Ave (region 1), is 2345, then $f_1^2 = 2345$. The level of neighbourhood is computed according to the distance between regions. In our simulation, density will affect the processes of founding and failure in the following manner: if the number of alive organizations in the neighbourhood of a specific firm belongs to the survival interval [DS_l ; DS_u], then the organization will have more probability to stay alive. Otherwise it will have more probability to die by the effect of "overcrowding" or "solitude", depending on whether the number is greater than DS_u or lower than DS_l (where DS_l and DS_u are respectively the lower and upper bounds of the density survival interval). The same idea applies to the process of founding where DB_l and DB_u respectively are the lower and upper bounds of the density founding interval. Moreover, considering what was said about the association between density and legitimation, DS_u may give us an idea about the maximum levels of legitimation for a particular type of industry.

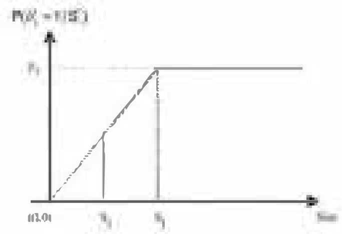
We will consider that age and size are also determinants of the firm's survival. Following Mata, Portugal and Guimarães (1995), size is an important determinant of the chances of survival and it exerts a negative effect on the instantaneous failure rate. So, we will consider that larger firms have more survival chances. Size will be measured by the number of employees. Also, experimental studies have shown that infant firms are more exposed to death (Carroll and Hannan, 1989). Thus, we will define S_l and A_l as the lower size and age thresholds above which an organization is more likely to survive. Therefore, assuming independence among density, size and age, the probability that a firm maintains in state 1, i.e., the survival for the firm i in the region j at time t, named P(Survival$_j$), can be formulated as :

$$P(Survival_j) = P(\delta_{ij}^t = 1 \mid D_j^t, S_i^t, A_i^t) =$$
$$= P(\delta_{ij}^t = 1 \mid D_j^t) \times P(\delta_{ij}^t = 1 \mid S_i^t) \times P(\delta_{ij}^t = 1 \mid A_i^t) \qquad (3)$$

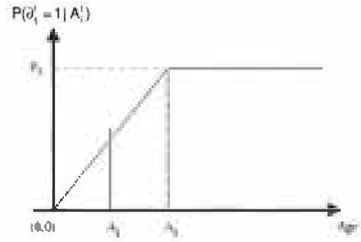
where S_i^t represents the size of the firm i at time t and A_i^t is the Age of the organization i (measured in years) at the time t. Figure 3 shows how the interval limits are associated with the probability of survival.



(a)



(b)



(c)

Figure 2: Probability of survival and interval limits for Density Survival (a), Size (b) and Age (c) and illustrative values for D_i, S_i and A_i. (at time t)

The parameters of the MAS (DS_l, DS_u, DB_l, DB_u, S_l, and A_l) need to be determined. They will be defined randomly at the beginning of the iterations but are optimized by calibration.

Preparation of the simulation

Real information is introduced in the simulation, in order to represent reality in a feasible way: distances between regions; a attraction level created by the authors, measuring the degree of industrialization and road access for industry location based on physical infrastructures, taxes, knowledge capital, inter-industry linkages, etc; number of firms in the textile industry as well as number of employees in that sector; resident population and vital (birth and death) rates for textile firms. The main data source for this work was the Firms and Plants Central File (FCEE) collected by INE - the Portuguese National Statistics Institute. We note that there is no information about density interval limits, and age and size thresholds. We prepared the simulation to generate the same conditions for birth and death of firms that exist in the four regions in the North-West of Portugal in the beginning of the study: the same number of firms in 1997 was distributed within the four regions according to the attraction levels; the age of each firm was set to 1 in first year; size was randomly generated, according to real size observed in each location.

To compare the impact of certain variables on the survival of firms, we used a Cox Proportional Hazard Model. Cox model (Cox, 1984) is used to determine the level of importance of some variables (X_{jt}) in the mortality of a cohort with time t. This model avoids the need to estimate ho(t), the formal distribution of the data (Cox, 1984, Mata, J. Portugal, P. and Guimarães, P., 1995).

$$h_j(t, X_{jt}) = h_0(t)\exp(X_{jt}\beta) \qquad (4)$$

Calibration of the final solution

The calibration process uses a Genetic Algorithm (GA). GAs can be seen as adaptive algorithms based on simulated evolution (Goldberg, 1989, Mitchell, 1997) using an appropriate metaphor of selection, crossover and mutation from Biology. In the context of the GA, a population of hypotheses is created in which each element of the population (hypothesis) contains the combination of parameters (DS_l, DS_u, DB_l, DB_u, S_l, and A_l) represented as genes. The GA then tries to search for the hypothesis that maximizes the fitness function value, by minimizing the distance between real and simulated output data. The output data are the simulated birth and death rates for the region of Ave. Fitness function contains a measure of the distance (Manhattan Distance) between the values of the vital rates produced from synthetic data and real observed data. We select half of the best fitted elements on each iteration and crossover carry out exchanging their genes. We also introduced a mutation rate of 10% in the selected population. We ran several arrangements of hypotheses and iterations of the whole algorithm, most of them containing 20 hypotheses and 10 iterations.

Simulation outline

The simulation algorithm considers the following steps:

1) Start-up: a fixed number of initial firms are randomly assigned to the regions taking into consideration the attraction level of each region.
2) Randomly define the parameters of the model: upper and lower boundaries for survival and founding, and lower limits for age and size.
3) Cycle for the MAS and hypotheses testing. Repeat until a predetermined threshold for the best fitted hypothesis is reached or a fixed number of iterations is attained (the one that happens first).

- *(if it is not the first iteration of the model): Apply Selection, Crossover and Mutation for the set of h hypotheses, and produce a new generation of h hypotheses. Runs the MAS so many times as the number of different hypotheses*
  - *Cycle for the generations of firms (using the parameters of a specific hypothesis):*
    - *(3.1) Produce N generations of firms; Determine in which 'regions' a new firm will be born; Determine which firms will die; Firms that survive can grow (increment their size);*
    - *(3.2) Compare vital rates (birth, death) produced by the population of firms from simulated data (at the region aggregate level) with those from real data and calculate the fitness function value to obtain the most fitted hypothesis.*

4) Return the best fitted hypothesis

Results and Validation

As we are firstly more concerned with the qualitative behaviour of the model, we started by observing the regularity of the simulated evolution of the number of firms in the four regions. Figure 3 shows the evolution of the population density in two of the regions for the textile industry. As the data from 1997 was the base for the simulation, we were curious about the evolution path of textile industry produced by the MAS. The figure shows simulated values for Ave (represented as "S Ave") and Minho-Lima ("S Minho-Lima)" in doted lines. Real values for those regions are represented in solid lines. Ave has the most part of textile industries in the region (as expected) and presents similar values as observed in reality. Even if the simulated evolution paths are not exactly the same as the real paths for every region, we believe that the MAS represents the reality in a coherent manner. Although this similarity may represent a validation of the simulation model in a certain way, we are more interested in analyzing further aspects of the survival of organizations that can be estimated from the simulation, rather than searching for a perfect fit between simulation and reality. In any case we need to be aware of the problem of overfitting the data.

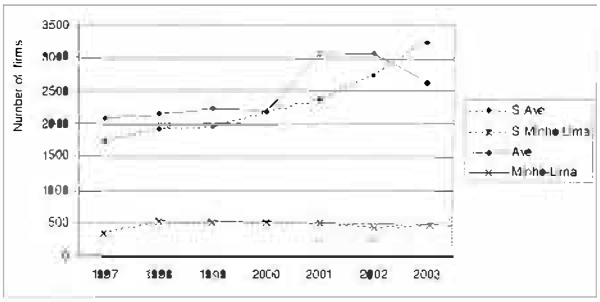


Figure 3: Real and simulated evolution of the number of textile firms in two regions of Portugal

We have also analyzed the Regression coefficients from a Cox proportional Hazard Model, estimate the impact of the covariates in the survival of organizations (for the Cox Regression, the cohort of firms that has been considered was produced by the most fit hypothesis). We have compared the sign of the coefficients with those obtained in other works (Mata and Portugal, 1994; Mata, Portugal and Guimarães, 1995; Carroll and Hannan 1989; Carroll and Hannan, 1992). The second column in Table 1 shows the predicted signs (obtained from literature - the same signs have been predicted in all studies) of the association between firm survival and the covariates.

Covariate	Predicted Sign	Coef	se(coef)	p
Contemporaneous Density	-	0.02688	0.00883	0.0023
Density at Founding	+	227.835	0.08537	0.0000
Size	+	-0.00412	0.00139	0.0031

Table 1: Coefficients of variables obtained from Cox Regression.

All the variables have a significant impact on the survival of firms, which can be seen from the p-values in last column (the standard error of the coefficients are represented as se(coef)). Carroll and Hannan (1989) explain that the density at founding has substantive implications on the survival of organizations. As we can see in Table 1, density at the time of a firm founding has a positive impact on the mortality of organizations while contemporaneous density (the number of firms existing at the moment of the death) has a corresponding negative effect. As expected, the effect of the size on firm survival is also negative, confirming what was said before and helping to validate the simulation model. Since the system seems to perform consistently when compared to reality (for comparable data), we tried to extract other useful information from the simulation using a calibration process. Our aim was to get an idea about the parameters for textile industry in the region of Ave. After running the simulation for a given number of iterations, the output values listed below correspond to the best combination of values (hypotheses), i.e. those which minimize the distance function used in this optimization process. These solutions of MASOS cannot be confirmed, because there is no available data for this kind of analysis. So a question arises whether this set of parameters is unique or, alternatively, whether there are more solutions. For that purpose we have analysed the outputs of several runs of the MAS simulation and retained the 10 best fitted hypotheses that constitute the solutions of the calibration process. We have found some small variations, as showed in Table 2.

DS_l	Dsu	DB_l	DB_u	S_l	A_l	Average Age
2	10	4	24	2	2	17.57
2	10	3	22	3	2	22.21
4	9	3	25	4	2	18.2
4	9	4	25	2	2	20.20
5	13	4	24	1	2	22.02
3	10	3	25	5	2	20.99
3	10	3	25	5	2	20.99

Table 2: Density interval limits (multiplied by 10) and Age and Size limits obtained from the best fitted solutions after GA calibration

We recall that DS_l and DS_u respectively are the lower and upper bounds of the density survival interval and DB_l and DB_u are respectively the lower and upper bounds of the founding (or birth) interval. A_l and S_l are the size, and age thresholds. These solutions define the hypothetical limits of the survival and founding intervals for the region of Ave between 1998 and 2003 (1997 was the starting year). Region of Ave seems to be very tolerant to newborn firms since estimated density birth interval limits lies between 0.3 and 2.5, which means that there is good receptivity for incoming firms in the period of analysis. New firms are almost always accepted but only some of them will survive: small size or infant firms are most likely to fail and older and larger firms will prevail. So, in the region of Ave, density survival (DS) levels of textile industries, measured in number of industries per square kilometre, stand more or less between 0.9 and 1.3 for the period of analysis. These values may constitute a type of legitimation limit that is implicitly established in the region. We have also obtained the ages and size limits (because small and younger firms are less likely to survive). We have found an age limit (A_l) of 2 and a size limit (S_l) that stands between 1 and 5.

CONCLUSIONS AND FUTURE WORK

Simulation with Multi-Agent Systems enables us to explore the situations where interactions are complex and nonlinear. In this work we focused our research on the density and studied its impact on the mortality of organizations using a Multi Agent based simulation. Data from the Portuguese textile industry have been used and we have tried to find out the density interval limits for this industry in the region of Ave, one of the most important textile clusters in Portugal. We have compared vital rates (birth and death rates) of simulated data with real data from Portuguese Textile Industry from 1997 to 2003. The real evolution path of the number of firms seemed quite similar when compared to the simulated evolution. Applying Survival Analysis, we could confirm the qualitative properties of the association among variables. Finally we have calibrated the MAS to identify the main determinants of organizational survival using a Genetic Algorithm and found the combination of these determinants that minimize the distance between real and simulated aggregate behaviour. The number of industries per square Kilometre stands more or less between 0.9 and 1.3 for the period of analysis We have also concluded that these firms are less likely to survive over the age of 2 and for a size interval that stands between 1 and 5.

In the future, we plan to compare the results with those obtained by some alternative techniques to provide a qualitative evaluation of the proposed model. Econometric techniques, such as regression analysis, for example, are often used to build up models where we can clearly distinguish between explanatory and dependent variables. We intend to compare tbe results of our method with classical techniques and conclude about the pertinence of our choice. Time series analysis is also suitable for modelling data, since we are going to generate sets of time indexed values. Another possibility could be Particle Swarm Optimization, particularly useful for our applications, as these are nonlinear optimization problems.

AUTHOR BIOGRAPHIES

Pedro Campos was born in Porto, Portugal, and obtained a degree in Applied Mathematics from Portucalense University and further post-graduate qualification (equivalent to M.Sc.) at the Faculty of Economics of the University of Porto (FEP). He is a Ph.D. Student in Managerial Sciences at FEP and member of LIACC - Laboratory of Artificial Intelligence and Computer Science.

Pavel Brazdil has obtained a Ph.D. in 1981 at the University of Edinburgh. He is a Full Professor at the Faculty of Economics of University of Porto and he is the coordinator of the group NIAAD of the Laboratory of Artificial Intelligence and Computer Science (LIACC).

REFERENCES

Albino, V., et al. 2003. "Coordination Mechanisms based on Cooperation and Competition within Industrial Districts: An Agent-based Computational Approach". Journal of Artificial Societies and Social Simulation, vol. 6, no. 4.

Carroll, G., Hannan, M. 1989. "Density Delay in the Evolution of Organizational Populations: A model and Five Empirical Tests", Admnistrative Science Quaterly, 34, 411-430.

Carroll, G., Hannan, M. 1992. "The Demography of Corporations and Industries". Princeton University Press, Pinceton, New Jersey

Cox, D.R. and Oakes, D. 1984, "Analysis of Survival Data". Chapman and Hall. Cambridge

Epstein, J., Axtell, R. 1996. "Growing Artificial Societies. Social Science from the bottom up". Brookings Institution Press, The MIT Press.

Goldberg, D. 1989. "Genetic Algorithms in Search. Optimization, and Machine Learning"

Hannan, M. and Freeman J. 1984, "Structural Inertia and Organizational Change". American Economic Review, 49. 149-164.

Lomi, A., Larsen, E. R. 1998. " Density Delay and Organizational Survival: Computational Models and empirical Comparisons". Computational & Mathematical Organization Theory, 3, 4. 219-247.

Mata, J. Portugal, P. and Guimarães, P. 1995. "The Survival of New plants: start-up conditions and post-entry evolution". International Journal of Industrial Organization. 13, 459-481.

Mitchell, T. 1997. "Machine Learning". Mc Graw Hill.

Stinchcombe, A. 1965. "Social structure and organizations". The handbook of Organizations, Chapter 4, March. J. (Eds.), Rand McNally & Company, 142-193.

Wissen, L. (2004). "A Spatial Interpretation of the Density Dependent Model in Industrial Demography". Small Business Economics, 22, 253-264.

Zhang, J. 2003 "Growing Silicon Valley on a Landscape: an agent-based approach to high-tech industrial clusters". Journal of Evolutionary Economics, 13, 529-548.

INDEX

SCS Publishing House