Project CASSIA
— Framework for Exhaustive and Large-scale Social Simulation —

Itsuki Noda, Yohsuke Murase, Nobuyasu Ito, Kiyoshi Izumi, Hiromitsu Hattori, Tomio Kamada, and Hideyuki Mizuta

1 Overview

Project CASSIA (Comprehensive Architecture of Social Simulation for Inclusive Analysis) aims to develop a framework to administer to execute large-scale multi-agent simulations exhaustively to analyze socially interactive systems. The framework will realize engineering environment to design and synthesize social systems like traffics, economy and politics.

The purpose of multi-agent social simulation is to provide tools to design social systems. It is impossible or quite difficult to carry out experiments of social phenomena in the real world by the similar way as experiments in physics or chemistry. Therefore, computational social simulations are indispensable means for social science.

Fortunately, progress of computational power has a potential to realize wider applications of computer simulation not only in physical phenomena but also in social problems. High performance computing (HPC) has been enabling several

Itsuki Noda
AIST, 1-1-1 Umezono, Tsukuba, Japan, e-mail: i.noda@aist.go.jp

Yohsuke Murase
Riken

Nobuyasu Ito
Univ. Tokyo & Riken

Kiyoshi Izumi
Univ. Tokyo

Hiromitsu Hattori
Ritsumei Univ.

Tomio Kamada
Kobe Univ.

Hideyuki Mizuta
IBM

simulation researches on large-scale weather, molecular dynamics, structures and architectures, and disasters. In addition to these physical phenomena, recently, social phenomena like economics, traffics, or information flow on networks attract many attentions as applications of HPC.

In order to make such social simulations on HPC available for wide-range users, the CASSIA framework consists of:

- MASS Planning Module: a manager module conducts effective execution plans of simulations among massive possible conditions according to available computer resources.
- MASS Parallel Middleware: an execution middleware provides functionality to realize distributed multi-agent simulation on many-core computers.

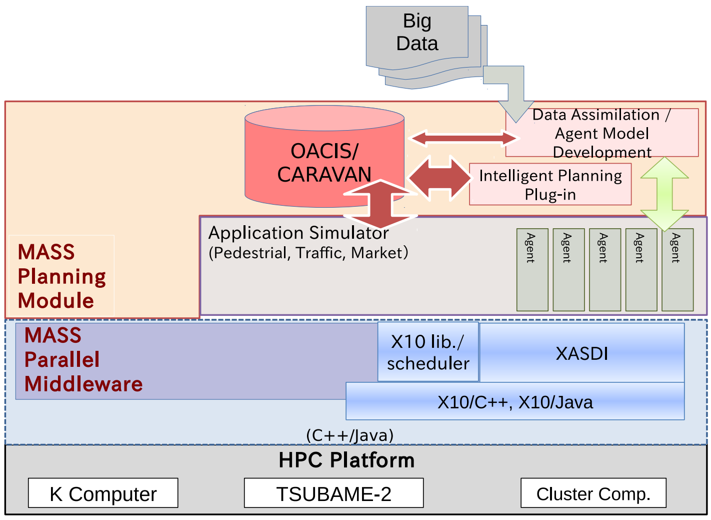

Fig. 1 Cassia Framework

2 MASS Planning Module

(Murase, Ito)

OACIS (Organizing Assistant for Comprehensive and Interactive Simulations) is a job management software for large-scale simulations. It controls a large number of simulation jobs executed in various remote servers, keeps these results in an organized way, and manages the analyses on these results.

CARAVAN provides more powerful scalability for exhaustive simulation. These functionalities are especially beneficial for the complex simulation models having many parameters for which a lot of parameter searches are required.

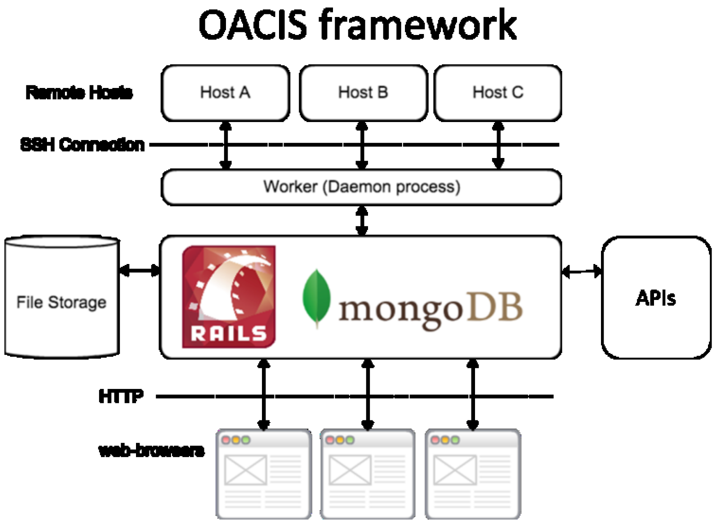

Fig. 2 OACIS

Fig. 3 CARAVAN

3 MASS Parallel Middleware

3.1 X10 Extentions and Plham

(Kamada)

Plham is a platform for large-scale and high-frequency artificial market simulation. It consists of models of markets for each stocks and three types of agents (high-freq. traders, short-term and long-term traders).

In order to enhance parallelism of computation, we introduce asynchronous computation in agents and communication between agents/markets, and provide high-level library to program them.

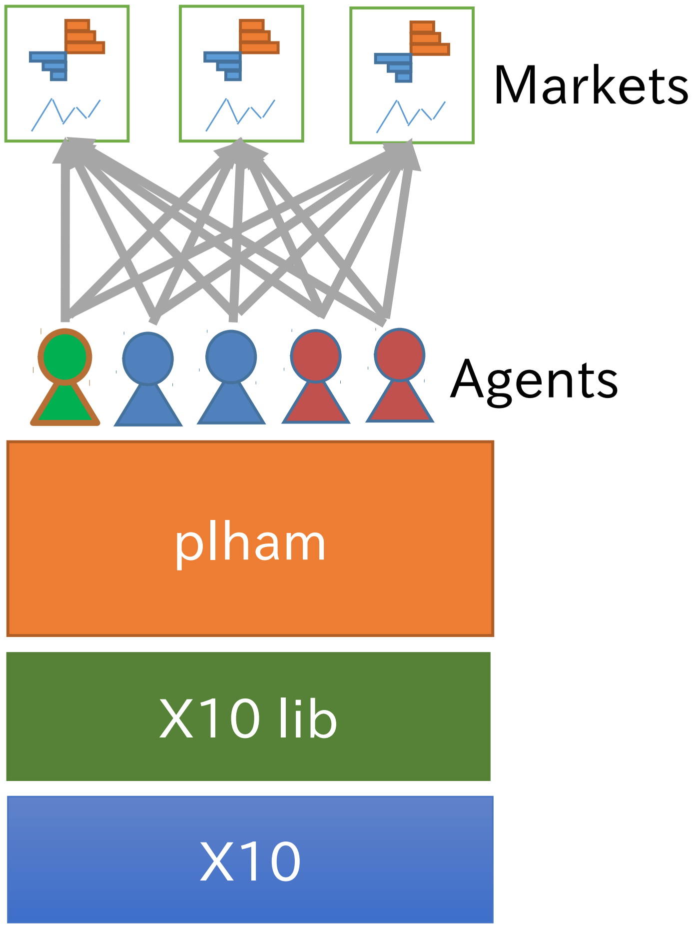

Fig. 4 Plham

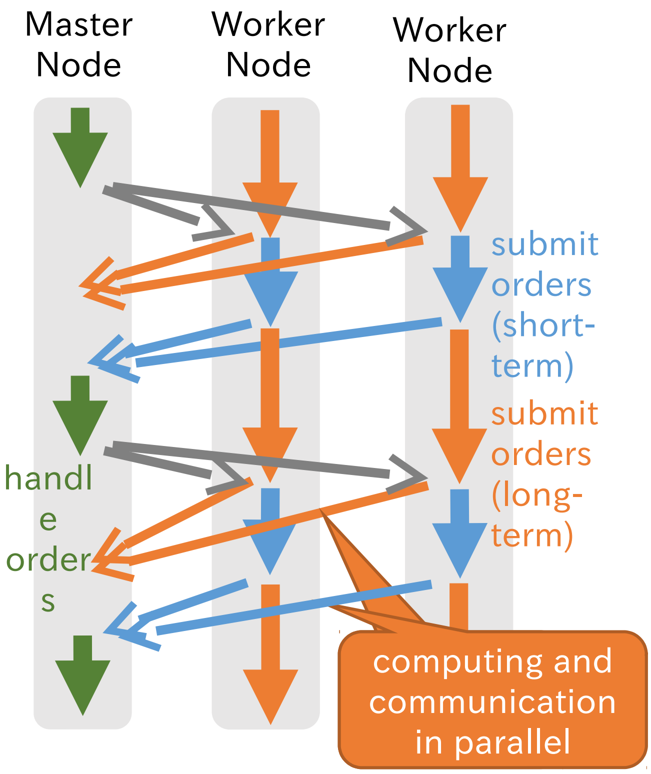

Fig. 5 Parallel Execution of Market Simulation

3.2 XASDI

(Mizuta)

XASDI is the Large-scale agent-based social simulation framework with billions of distributed agents that provides easy-to-use API bridge with Java and X10-based runtime for high scalability. XASDI environment executes various social simulations written in Java with distributed agents and managers written in X10.

4 Applications

4.1 Market Simulation

(Izumi)

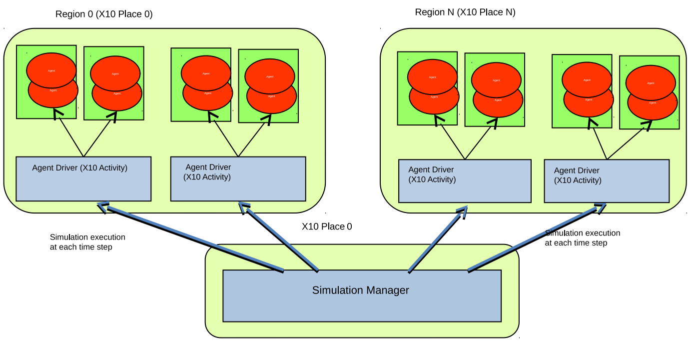

Fig. 6 XASDI

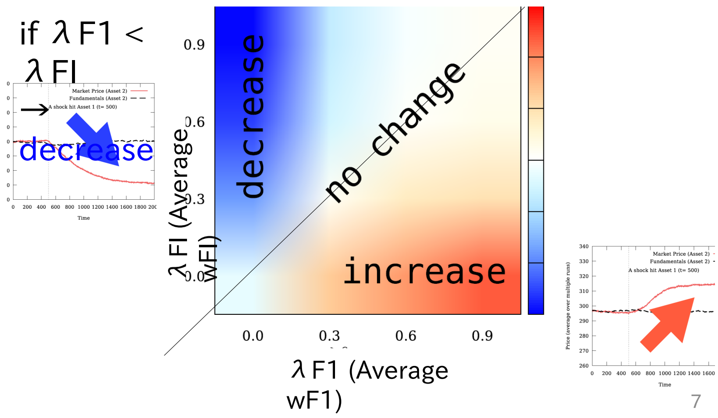

Fig. 7 Phase Diagram of Market Simulation

4.2 *Pedestrian Simulation*

CASSIA Framework can illustrate a trade-off structure of social problems. The most of social problems like planning of evacuations from disasters are not simple optimization problems but dilemmas among multiple objective functions. We will show an example to apply CASSIA Framework to find such trade-off structures using evacuation planning for Nishiyogogawa-ku, Osaka, which includes over 300 control parameters[5]. Because of the large degrees of freedom, the search space of this problem is so huge that the solution of this problem require high-performance computing like K-computers.

4.2.1 Pedestrian simulator: CrowdWalk

In order to simulate evacuation, we employ simulator CrowdWalk [11, 10]. CrowdWalk is a pedestrian simulator that limits human movement to one-dimensional movement on the road network. Road network is composed of nodes and links, and CrowdWalk controls large number of pedestrian movements on wide area at high speed.

We use a road map of Nishiyodogawa-ku, which consists of 7,624 nodes (crossings) and 10,707 road links (figure 8). The local government select 86 official refuges for this area. We suppose that the population of this area is 54,909 who are distributed in 146 small zones in the area. We also suppose that all people in a zone follows an evacuation rule that asks people to go to a certain destination with one via point in a map. The destination and the via point is selected from the 86 official refuges and from 533 candidates for via points, respectively.

Fig. 8 Nishiyodogawa Area (left) and Road Map (right) used in Pedestrian Simulation

4.2.2 Multi-Objective Optimization

As described before, evacuation planning is a dilemma between evacuation time and simple-ness of evacuation rule. From the viewpoint of the minimization of evacuation time, it is better to use the result of mathematical optimization like maximum network-flow [3]. However, we need to guide large number of people that include ones who are not acquainted with the place like visitors. So, the guidance should be simple enough to understand and to follow easily.

In order to know the relationship of these two objectives, we apply NSGA-II(Elitist Non-Dominated Sorting Genetic Algorithm) [2], a genetic algorithm for multiple objective optimization[1].

For the first objection function, the evacuation time, is estimated by simulation using CrowdWalk for each guidance plan.

For the second objective function, the simple-ness of evacuation plans, we introduce 'entropy' of the plan as follows. The basic concept of the entropy is illustrated in figure 9. Suppose two connecting zones, z_i and z_j in the area, and populations of the zones are n_i and n_j, respectively. If the two rules for these zones has same via point and destination, then the entropy is zero. Otherwise, the entropy of this pair is defined by:

$$H(z_i, z_j) = -(n_i/(n_i + n_j)) \log(n_i/(n_i + n_j)) - (n_j/(n_i + n_j)) \log(n_j/(n_i + n_j)).$$

Finally, we use total entropy $H = \sum_{z_i, z_j} H(z_i, z_j)$ for the index of the complexity of the guidance (negative value of the simple-ness).

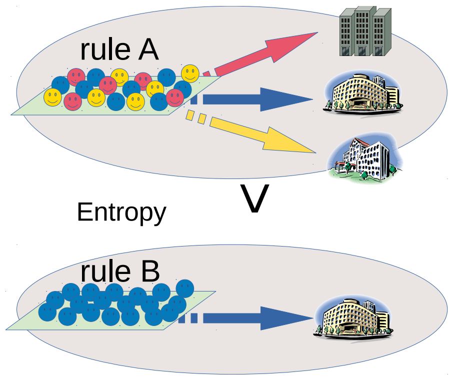

Fig. 9 Rule Entropy

4.2.3 Experimental Result and Discussion

In order to run NSGA-II for this evacuation plan, we utilized OACIS[12] to manage the large number of runs. The search space of this problem is so huge ($R^{73} \times 533^{146} \times 86^{146}$) and NSGA-II requires large number of populations (about 100–1,000). So we need to run so many runs for the optimization. In the experiment, we runs 500 generations with 100 population for the optimization, which means we run 500,000 simulations [1] for this experiment.

To control NSGA-II procedure, we utilize the ruby plug-in facility of OACIS. In the actual GA procedure, we use 'simulated binary crossover' and 'polynomial mutation' for creating new generations, and Paleto ranking mechanism to determine the selection.

Figure 10 shows the result of the experiment. In this graph, vertical and horizontal axes indicate evacuation time and the complexity of plan (total entropy scaled

[1] We runs 10 simulations for one evacuation plan to calculate average evacuation time.

by 100), respectively. The color of the dot indicates the generations. From this result, we can see that the evacuation plans are improved by progress of generations, and almost saturate to boundary of 3000 for evacuation time and 2100 for complexity of guidance. In order to minimize the evacuation time, we need to choose relatively complex guidance (the complexity is about 2200 rather than 2100). On the other hand, if we consider simplify the complexity, the evacuation time increase drastically up to 7000. And, we can see reasonable guidance will exist the most left-bottom area of the Paleto front in this graph.

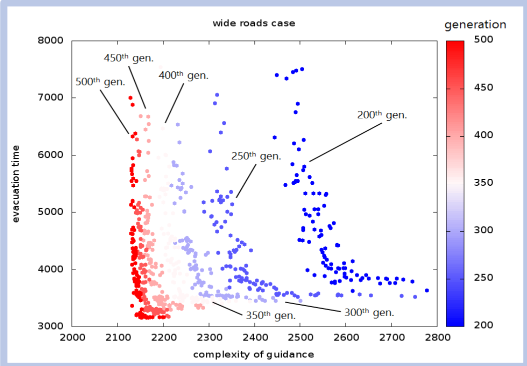

Fig. 10 Result of Evacuation Simulation (wide road)

Figure 11 also shows the result of the case that the people use only pedestrian road. In this case, the boundary of the evacuation time increase to 4500, but the complexity of guidance is similar to the previous case.

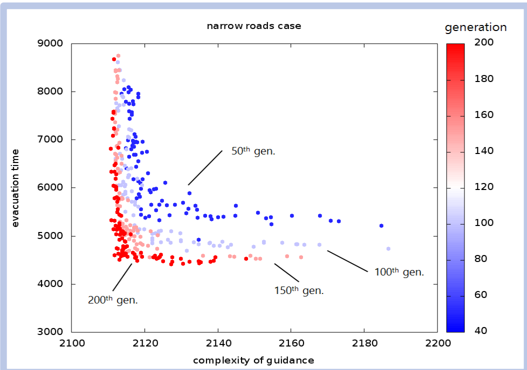

Fig. 11 Result of Evacuation Simulation (narrow road)

As shown in this section, evolutionary methods using OACIS can be a powerful tool to investigate such social problems. We also succeed to apply CARAVAN to run the same procedure on K-computer. This combination enables to run larger scale of simulations and search spaces.

4.3 Traffic Simulation

(Hattori, Ito?)

5 Computational Roadmaps of Social Simulations and Future

As described above, our purpose is to determine how HPC contributes to the advancement of research on social simulation or to clarify the computational power required for real applications of social simulation. In this section, we focus on three applications and try to develop roadmaps for them.

In the development of these roadmaps, we adopted two indexes to measure the computational cost, "number of situations" and "complexity of one simulation session". We considered exhaustive evaluation by simulation as a key methodology of social simulation. Therefore, to evaluate the model, examining many conditions and models is important. The index of "number of situations" indicates this number. Meanwhile, ordinal computational cost of a simulation, which is determined by the number of entities and the number of interactions among the entities, is important. In addition, in multiagent simulation, the computational cost of thinking of each agent is significant. In the following discussion, we integrate these complexities as "complexity of one simulation session".

5.1 Evacuation/Pedestrian Simulation

The main target of evacuation simulation is not to find an optimal plan of evacuation for a given disaster situation, but to evaluate the feasibility and robustness of executable candidates of evacuation plans or guidance policies. Because of natures of disasters, it is difficult to acquire complete information to determine the conditions of evacuations in the event of a disaster. Therefore, it is almost impossible to validate optimality for each disaster. Instead, local governments should strive to establish feasible plans that will work robustly in most situations of disasters. This means evaluation of evacuation plans should be done under widely varying disaster scenarios. A massively parallel computer simulation will make such evaluations easy and effective.

Several simulations have been performed for evaluating such evacuation plans [14][15][7][13]. For example, a simulation of an evacuation from a Tsunami struck city in Tokai area in Japan was performed, where a massive damage is expected to occur due to the great Tokai-Tonankai earthquake. To help understand the importance of the relationship between evacuation scale (populations of evacuees) and effectiveness of evacuation plans, we conducted the following exhaustive simulations considering various sizes and evacuation policies (evacuee's origin-destination (OD) plans). The simulation results indicate that the scale of evacuation can be grouped into two categories, namely, "large" ($>$ 3,000 evacuees) and "small" ($<$ 3,000 evacuees). Each evacuation plan has similar relative effectiveness in each category. The actual evacuation size (population) may change based on various factors such as daytime/nighttime, number of visitors/travelers, weather, and special events. This implies that citizens and local governments should consider at least two plans for large- and small-scale evacuations.

We execute the evacuation simulation described above to arrive at a reference point for illustrating computational costs of various actual applications. In the above simulation, we considered the following scenarios:

- 2,187 OD plans and
- 8 cases of evacuation population (70–10,000 agents).

Therefore, in total, 17,497 simulation scenarios were executed over about 30 days when using a single process on Xeon E5 CPU (2.7 GHz). We denote this reference point as the rectangle "city zone, TSUNAMI" in figure 12.

We can easily extend the simulation scale. Although a population of only 10,000 is considered in "city zone, TSUNAMI", we can extend the simulation to a more densely populated area such as in Tokyo. For example, we performed a similar simulation analysis in the Kanazawa area, which is located on the coast along the Japan Sea and experiences snowfall in the winter. In this case, the population size is similar (about 6,000 agents), but the number of combinations of scenarios increases to 4,194,304 (2^{22}). The rectangle "city zone, TSUNAMI and HEAVY SNOW" in figure 12 denotes this calculation cost.

We can further extend the simulation to a large scale with a larger number of scenarios. Kitasenju area, a large transfer station surrounded by rivers, has a population of 70,000, and the computational cost of simulating this area is denoted by "dense-population zone, complex disaster" in figure 12. Because this area is densely populated and complex, we have combinations of 44 policy candidates, that is 2^{44} scenarios. In the case of Tokyo, we need additional computational power. In figure 12, "megacity" corresponds a huge city such as Tokyo. In this case, the size of evacuation and the number of possible scenarios is very large. Therefore, peta- or exa-scale HPC is required to handle such simulations.

5.2 Traffic Simulation

Road traffic is an important domain from the viewpoint of applying social simulation. Traffic simulation has been extensively researched over a long period, and recently, the focus has been on multiagent simulation, in which each agent behaves

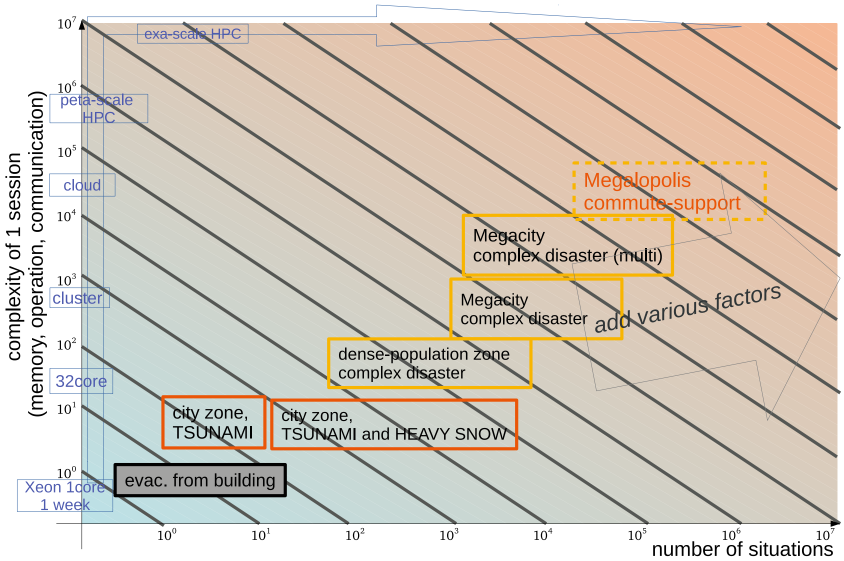

Fig. 12 Roadmap of Evacuation Simulation

according to its own preferences and inference rules. Big data advances in computational power enable us to perform such detailed simulations.

[8] have been developing a traffic simulator called IBM Mega Traffic Simulator that can run large scale traffic simulations on XASDI middleware. The main feature of this simulator is its ability to reflect individual drivers' preferences. Using this feature, according to big-data, we can adapt parameters in the simulation that cause differences in drivers' tendency.

To create a reference point for the roadmap of the traffic simulation, we considered the case of evaluating road restriction policies for road construction in the Hiroshima area[9]. In this case, we performed simulations of the following scales:

- 70,000 agents (trips), 120,000 road links, and 15 hours and
- 20 cases

In this case, the calculation required about one day when using a single process on Xion E5 CPU. We denote this reference point as "million city, road plan" in figure 13.

We can draw out the roadmap from this reference point. When considering the Tokyo area, the number of agents increases up to about 2 million and the number of road links increases to about 610,000. Moreover, if we consider a larger area such as the Tokyo metropolitan area, the population increases about 4 million and the number of road links increases to 2.5 million. These calculation costs are plotted as "Tokyo, traffic control" and "metropolis, traffic control" in figure 13.

When we consider a big event, we must list a large number of cases to evaluate the robustness of road traffic to accidents, whereas the scenarios mentioned above pertain to normal situations that are repeated every day. Because various situations affect traffics, the number of situations increases quickly. These costs are plotted as “Tokyo, big event”, “metropolis, big event” and “whole Japan, big event” in figure 13, and they require exa-scale computational power.

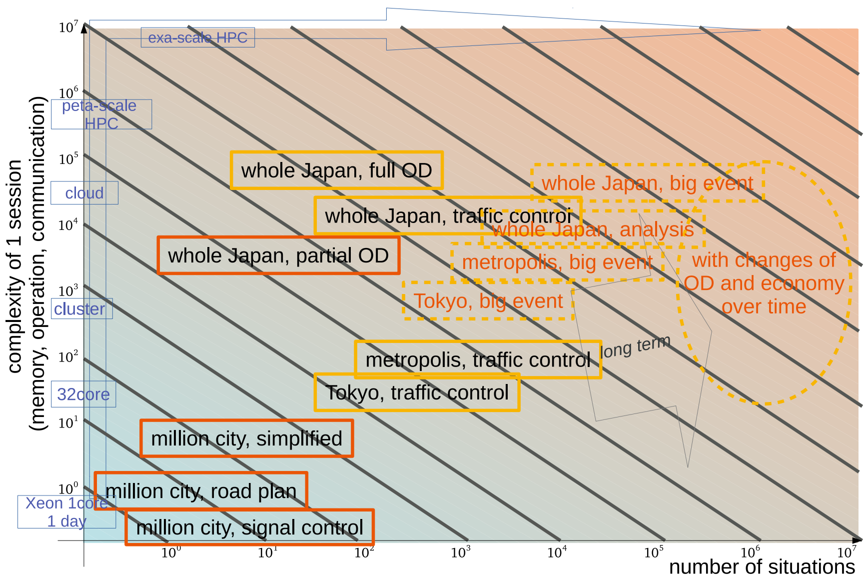

Fig. 13 Roadmap of Traffic Simulation

5.3 Market Simulation

Market simulations are another important application of multiagent simulations, in which agents directly affect each other by selling/buying stocks and/or currencies [4]. Compared with evacuation and traffic simulations, market simulations are not constrained by physical space. Therefore, the time cycles of agents’ interactions may be quite short. Moreover, the ways of thinking of agents show large variations. This means that the market simulations also require huge computational cost.

As the reference point of the calculation cost in market simulations, we present the case of “tic size” evaluation. In this scenario, we conducted a simulation of multiple markets having different tic sizes, which is the minimum price unit for trading stocks. Market companies such as Japan Exchange Group internationally compete with each other by providing attractive services to traders. A small tic size is one of such services that considerably increases cost. Therefore, such organizations need evaluations of changes to such services in advance. In collaborative works with

Japan Exchange Group, we conducted a simulation experiment to find key conditions that determine market share among markets. In the simulation, we considered the following scenario:

- one good in two markets, 1,000 agents, and 10 million cycles
- five cases of tic size and 100 simulation runs per case

This simulation takes about one day when using a single thread on a Xeon E5 CPU. As the reference point, we plot this as “tic size” in figure 14.

We are considering extending the market simulations to various applications used for stock market analyses. For example, it is in the interest of market companies to determine “daily limit” and “cut-off” prices[6]. In this case, the simulation must handle 10–20 goods. Moreover, evaluating the effects of “arbitrage” [4], which involves trading rather quickly in intervals of milliseconds, is important from the viewpoint of maintaining sound market conditions. This will increase the computational cost, as plotted in figure 14. Another topic is the evaluation of “Basel Capital Accords”, which deal with the soundness of banks in markets. In the present study, we executed the case of three names for the Basel Accords, but we will extend it to 100 names in the real application.

The evaluation of “systemic risks of inter-bank network” is an important issue in market evaluation. However, currently, the computational cost of a naive simulation exceeds exa-scale HPC.

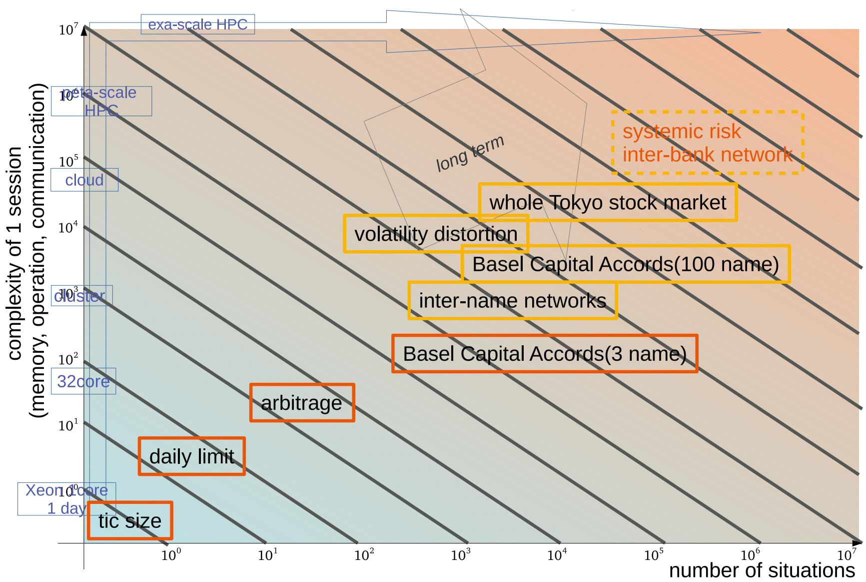

Fig. 14 Roadmap of Market Simulation

5.4 Future Issues

In the roadmaps, we count the number of scenarios naively, because we need to know the actual numbers of the possible scenarios. We can apply several methods based on design of experiments and other optimization/learning methods to reduce the number of scenarios we should run. OACIS and CARAVAN have abilities to introduce such functionality.

We also need to investigate the cost of thinking part of each agent. In the evaluation above, we assume that the intelligence of each agent will not change, so that the complexity of the thinking in each agent is constant. But, for further simulation researches, we need to introduce more sophisticated and complex thinking engine to realize more intelligent and adaptive behaviors like human. This is still open issues.

The multiagent social simulation is evolving research domain and still under-establishing phase. However, requests from application fields become stronger and stronger. So, it is important to determine a measure to know achievements will be important. The roadmaps shown in this article will beocme a testbed to provide such measures. Also, the CASSIA framework shown in this chapter will provide powerful tool to push forward the research on this roadmaps.

References

1. K., D.: Multiobjective Optimization using Evolutionary Algorithms, vol. 16. John Wiley and Sons (2001)
2. K., D., S., A., A., P., T., M.: A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii. In: S.M. et al (ed.) PPSN VI Proceedings of the 6th International Conference on Parallel Problem Solving from Nature, *Lecture Notes in Computer Science*, vol. 1917, pp. 849–858. Springer, Berlin, Heidelberg (2000). DOI 10.1007/3-540-45356-3_83
3. K., K., R., N., Y., Y., K., F.: Experimental analyses of the evacuation planning model using lexicographically quickest flow [in japanese]. Transactions of the Operations Research Society of Japan **59**, 86–105 (2016). DOI 10.15807/torsj.59.86
4. Kawakubo, S., Izumi, K., Yoshimura, S.: Analysis of an option market dynamics based on a heterogeneous agent model. Int. J. Intell. Syst. Acc. Fin. Mgmt. **21**, 105–128 (2014). DOI 10.1002/isaf.1353
5. Matsushima, H., Noda, I.: Analysis of trade-off in evacuation plan using evolutionarily exhaustive simulation. In: Proc. of the 23rd International Symposium on Artificial Life and Robotics (AROB 23rd 2018), pp. OS15–4. International Society of Artificial Life and Robotics (2018)
6. Mizuta, T., Izumi, K., Yagi, I., Yoshimura, S.: Design of financial market regulations against large price fluctuations using by artificial market simulations. Journal of Mathematical Finance **3**(2A), 15–22 (2013). DOI 10.4236/jmf.2013.32A003
7. Nurdin, Y., Yuliana, D.K., Noda, I., Soeda, S., Yamashita, T.: Disaster evacuation simulation with multi-agent system approach using netmas for contingency planning (meulaboh case study). In: Proc. of 5th Annual International Workshop & Expo on Sumatra Tsunami Disaster & Recovery 2010, p. ??? AIWEST-2010 (2010)
8. Osogami, T., Imamichi, T., Mizuta, H., Morimura, T., Raymond, R., Suzumura, T., Takahashi, R., Ide, T.: Ibm mega traffic simulator. IBM Research and Development Journal (2013)

9. Osogami, T., Mizuta, H., Ide, T.: Identifying the optimal road closure with simulation. In: Proceedings of the 20th ITS World Congress Tokyo 2013 (2013)
10. T., Y., S., S., M., O., I., N.: Development and application of high-speed evacuation simulator with one-dimensional pedestrian model. Jpournal of Information Processing Society Japan **53**(7), 1732–1744 (2012)
11. T., Y., T., O., I., N.: Implementation of simulation environment for exhaustive analysis of huge-scale pedestrian flow. SICE JCMSI **6**(2), 137–146 (2013)
12. Y., M., T., U., N., I.: A tool for parameter-space explorations. Phyisics Procedia **59**, 73–76 (2014). DOI 10.1016/j.phpro.2014.08.134
13. Yamashita, T., Matsushima, H., Noda, I.: Exhaustive analysis with a pedestrian simulation environment for assistant of evacuation planning. In: Prof. of PED 2014, pp. SE05–3 (2014). DOI https://doi.org/10.1016/j.trpro.2014.09.047
14. Yamashita, T., Soeda, S., Noda, I.: Evacuation planning assist system with network model-based pedestrian simulator. In: J.J. Yang, M. Yokoo, Z.J. Takayuki Ito, P. Scerri (eds.) Principles of Practice in Multi-Agent Systems (Proc. of 12th International Conference, PRIMA 2009), pp. 649–656. Springer, Heidelberg (2009)
15. Yamashita, T., Soeda, S., Noda, I.: Assistance of evacuation planning with high-speed network model-based pedestrian simulator. In: Proceedings of Fifth International Conference on Pedestrian and Evacuation Dynamics (PED 2010), p. 58. PED 2010 (2010)