EXPLORING an EPIDEMIC in an E-Science Environment

Rules of bio-epidemic and e-epidemic inspire scientists to create a live, scalable interconnected environment for effectively managing situations in nature, society, and the digital virtual world.

By HAI ZHUGE

The first outbreak of the Severe Acute Respiratory Syndrome (SARS) epidemic in China in 2003 was a great global tragedy. The loss of lives was devastating, propelled in part by a total lack of knowledge for taking proper control measures, for sharing vital information, and for determining and disseminating effective medical treatment in a timely and perhaps life-saving manner [8]. The China Knowledge Grid Research Group explored the spread rules of SARS from the initial stage of the epidemic in an effort to build a cross-regional cooperative research and management environment for public health.

To design a simulation of the SARS epidemic, we generalize the society to be investigated as a $n \times n$ grid, where n is an integer representing the scale and capacity of the society. It will be initialized before simulation. Each square unit has the following six possible colors:

- Gray representing healthy people;
- Red representing infected people;
- Yellow representing people who have come in close contact with infected people;
- Pink representing people suspected of carrying the infection as they have been in contact with infected people and present SARS-like symptoms;

The China Knowledge Grid Research Group *explored the*
spread rules of SARS from the initial stage of the epidemic in an
effort to build a cross-regional cooperative research and
management environment for public health.

• Green representing people who have recovered and now have the antibody; and,
• Black representing people who have died from SARS.

With the exception of the black squares, the colored squares move around randomly at a certain radium to simulate the dynamic characteristic of human behavior. The gray squares turn yellow once they contact the red squares. The yellow rectangles randomly turn red and pink at certain ratio. The yellow and pink squares will turn gray (meaning the person is not infected) if they have not turned red after a preset latent period. The red squares turn black and green at the preset death rate and recovery rate, respectively. Figure 1(a) shows the state transformation graph of each square. The healthy state, immunized state, and dead state are stable. The latent period, death rate and recover rate are preset before simulation with reference to statistical data of the SARS epidemic and infection features.

The infectious ratio can be changed through adjusting the proceeding time unit during simulation. The isolation control of the infected person is simulated by allowing the number of gray rectangles to be the neighbor of red, yellow, and pink squares respectively.

The simulation runs across platforms and the process can be observed from any Web browser. The main data structure includes two arrays: one records the current state of each person and his/her current location, and the other records the current occupation status of the two-dimensional grid. The program processes the simulation and updates the display at the given time unit to simulate a common daily situation. Figure 1(b) shows the simulation interface viewable from a Web browser.

SIMULATION METHOD AND RESULTS

Change the population density to observe the effect of the epidemic situation. The simulation results show that the number of infected people is sensitive to the population density, and the peak region of a low-density population generally occurs later than the peak of a high-density population.

Change the time for isolation control to observe the effect of epidemic situation. The simulation results show that the isolation control measure (isolating only the infected person) is effective before the peak comes. The isolation measure can curb the height of the peak, but has almost no effect on the duration of the epidemic situation. The peak region under isolation control is typically earlier than that without control. Figure 2 shows four sets of simulation results that compare the situations under the control measures taken at different stages and different population den-
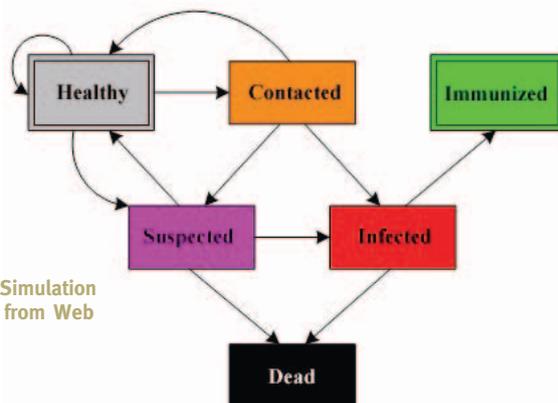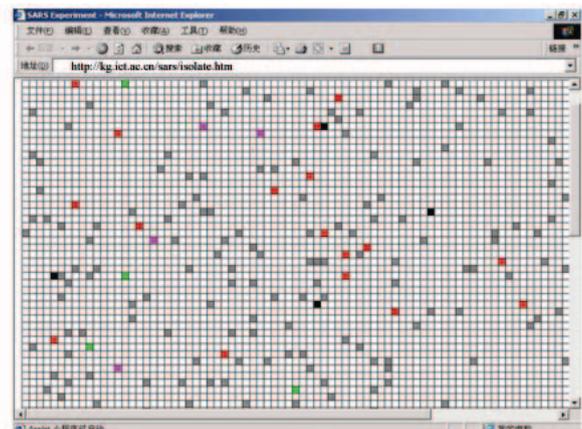

Figure 1. Simulation observed from Web browser.

(a) State transformation of each colored square.
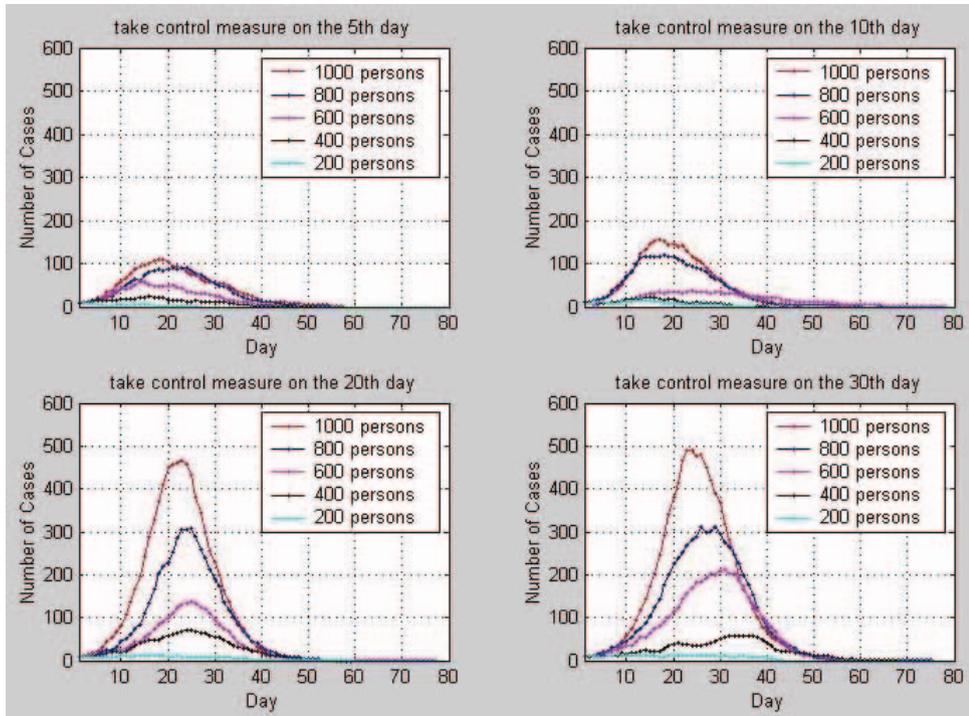
(b) Visualized simulation interface.

Figure 2. Compare the situations under the control measures of different stages and different populations.

The simulation of the epidemic situation of a city, a country, or a worldwide region requires the co-simulation of an epidemic situation of geographically dispersed regions or communities. Co-simulation can reflect a general epidemic situation and the impact of one on another. Figure 3(a) shows the interface of the co-simulation of two different regions *A* and *B*, where the path between the two regions may transport the infected, contacted, or suspected person. Assume that SARS first spread in region *A*; we can observe the impact of the traffic flow on the epidemic situations in the two regions by adjusting the scale of flow. Figure 3(b) shows the total "infected" number is sensitive to the scale of flow: the greater flow causes the larger number of infected people (as the pink and black curves show), but after a certain period of time (say, 30 days) the number is not sensitive to the scale of flow. Figures 3(c) and (d) show the impact of the scale of flow on these two regions. They show that we can totally control the epidemic situation spread from *A* to *B* when reducing the scale of flow to a certain level. Therefore, it is not necessary to completely break the traffic flow between regions.

The simulation system enables users to arbitrarily select regions from a list to conduct co-simulation of multiple regions. The implementation includes the following steps:

- Integrating the displays of the simulation of different regions to obtain an overall view;
- Integrating internal data structures;
- Setting constraints of the border (traffic flow) between regions; and
- Executing simulation according to the integrated data structure and constraints.

IMPLICATION
Simulation under different conditions indicates:

sities. The higher population density is more sensitive to the isolation control measure.

Change the isolation extent to observe the epidemic situation. The isolation control measure has different extents: isolate the infected people, isolate the suspected people, isolate the contacted people, and isolate the people on possible infection chains. We observed the number of infected people by increasing the number of controlled objects on the contact chain. The experiment shows the epidemic situation is not sensitive to the extent of isolation. This is because the infected person is not so contagious in the latent period, so only a very small number of people are in the contact region in that phase of infection.

The active degree of people participating in social activities reflects the psychological factor. We first changed the move radium of activity to watch the effect of an epidemic situation under different isolation measures, and then changed the move frequency to observe the effect. The simulation results show the number of infected people is not sensitive to the move radium, and the middle move-radium leads to a bit larger number of infected people than does the big and small move radium. The simulation also shows that the number of infected people is sensitive to the move frequency. We randomly identify the rectangles as groups and assign them different move frequencies to stand for different behavior features of different communities in the society. The group with the higher move frequency leads to a larger number of infected people than does the group with lower move frequency.

- The isolation control measure is only effective at the early stage of an epidemic. Therefore, establishing precautionary systems can help us determine the epidemic situation early and use control measures promptly.
- Isolating a potentially infected person on the contact chain does not have an obvious effect on controlling the epidemic situation; therefore, only an infected person requires isolation. The isolation measure and personal information collection requires relevant laws—a reflection of the social aspects of the simulation.
- The epidemic situation is sensitive to population density. Thus, raising the protection standard in densely populated locations, especially hospitals, can help control an epidemic situation.
- Co-simulation shows that the epidemic situation is sensitive to the scale of flow between regions and can be effectively controlled from affecting each other by reducing the scale. Therefore, absolute isolation between regions is not necessary.
- The public psychological response to an epidemic situation will reduce the contact frequency between people and public gatherings. This is a kind of natural resilience of a healthy species that can prevent the epidemic situation from getting endlessly worse. But natural psychological response depends on the distribution of timely epidemic information.

Evaluations taken during the second SARS occurrence in April 2004 in China verify these results.

ANALYZING AND MANAGING AN EPIDEMIC IN A DYNAMIC SMALL WORLD

Simulation shows the "infected" number is sensitive to population density. A survey of real cases in China also shows that SARS mainly spreads in small, close contact groups or cliques. A survey conducted after China's first outbreak shows that the largest close contact infection tree only includes 37 people. Its second occurrence in China indicated the largest spread chain had only five people. This characteristic makes it feasible to analyze the spread network of SARS.
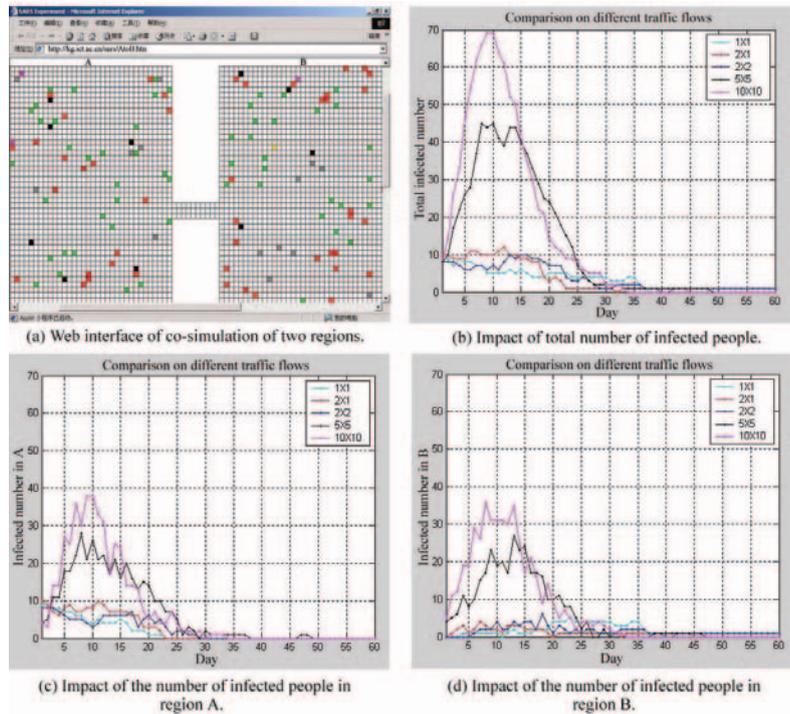

(a) Web interface of co-simulation of two regions.

(b) Impact of total number of infected people.

(c) Impact of the number of infected people in region A.

(d) Impact of the number of infected people in region B.

Figure 3. Co-simulation of two regions.

The spread network is live and scalable because nodes have state and life spans; the dead nodes will be removed and new nodes will be added from time to time. By zooming out we can see cliques—the small close-contact spread networks. Nodes within the same clique share some common social roles (for example, nurse and patient) and they have a higher probability to be mutually infected than the nodes belong to different cliques.

People will act intelligently to avoid infection if they are armed with protective information and understand the current situation (that is, the general epidemic situation and the status of their surrounding environment). The infection model includes feedback from the environment, along with the following five variables: The *contact frequency* between nodes. The node with more and higher contact frequency arcs has higher *contact rank*. The *infected ratio* reflects the probability of a node to be infected. The *infect ratio* reflects the probability of a node infecting other nodes. The infected ratio of a node is a function of the infected ratio of all the direct-linked nodes and the contact frequencies of these nodes. The *infect rank* of a node is a function of its contact rank and infected ratio. The social feedback mechanism enables the general contact frequency to be the negative function of the total infected number.

The number of infected people is a function of time. Peak usually occurs within the first month. A survey in Beijing shows that infected people are not as

Various sensors and sensor networks, mobile digital devices,
and robots extend the Internet to a pervasive interconnection environment,
which can automatically monitor and collect societal information to form an
ideal environment for simulation.

contagious in the latent period, therefore, the five variables are functions of time and we can filter out different granularities in spread network cliques by determining different effective contact frequency. On the other hand, people move to play different roles in the society, and a role can be involved in nodes of different cliques.

A live scalable spread network model is a function of time and the cooperation of the following three layers: A scalable live network; an infection model; and, a role management model. Previous research isolates the network model, infection model, and behavior.

The power-law distribution of the self-organized network [3, 6, 7] and the dynamic characteristic of the spread network require effective epidemic control measure to prevent the high rank nodes in the spread network from being infected. Persons and roles involved in the high rank nodes should be monitored regularly.

Exploring Epidemic with the Knowledge Grid

Current Grid and P2P computing are still limited in their semantic ability to support intelligent applications [2, 5]. Toward the future interconnection environment [12], China's e-science Knowledge Grid environment is to gather epidemic researchers, IT professionals who develop and maintain the environment, people in different health status, and policymakers to build mutual understanding and cooperation, and promote effective information sharing through normally organizing, semantically interconnecting, and dynamically clustering digital resources on dynamic, large-scale, scalable, and semantic-rich networks.

Various sensors and sensor networks, mobile digital devices, and robots extend the Internet to a pervasive interconnection environment, which can automatically monitor and collect societal information (for example, people's temperature and movement at the entrance of airports, railway stations, and theaters) and nature (for example, temperature, humidity, pollution of water), to form an ideal environment for simulation, real-time study, and exploration of the intrinsic relationship among epidemics, society, and nature.

The synergy of content retrieval, extraction, fusion, filtering, and data mining technologies supports automatic discovery of different types of contact relationships among self-organized individuals in the environment, by searching and analyzing the evolution of user logs, interests, email [9], and online digital records (personal information, family members, colleagues, and employers). With interconnection semantics, the e-science Knowledge Grid environment can effectively collect, manage, and update personal information, to monitor and update the epidemic situation, and to make spread analysis and prediction.

Information redundancy and inconsistency occur across isolated information systems when a patient travels between regions and sees doctors in different hospitals. A national or worldwide epidemic e-science environment can accurately collect and manage cross-region or cross-country epidemic information, eliminate inconsistency and redundancy, share resources, analyze epidemic situation, and carry out cooperative research by integrating multiple regional systems. Versatile information sources should be organized in semantic normal forms under integrity constraints to guarantee correct information retrieval and update. Information distributed in different regions can be easily integrated by using join and merge operations based on the Resource Space Model for organizing and managing resources within the Knowledge Grid [11].

E-Epidemic and Flows

The e-science Knowledge Grid environment not only supports cooperative simulation and management of epidemics, but also the study of its e-epidemics. With human intelligence, e-virus programs can even behave actively and intelligently to maximize the damage of the network by attacking the high rank nodes. Different types of e-viruses use different epidemic models similar to that in biological epidemiology, thus different immune strategies are needed to conquer them [1, 4]. A combined preventive strategy is necessary to defeat the complex attack. The study of different types of biological epidemics helps control an e-epidemic. The experience

The practice of simulating the SARS epidemic provides us the ability to experience the overall evolution processes of a live scalable network.

and three-layer model of an epidemic spread can apply to effective control of harmful information spread and widen the useful information spread.

The study of an e-epidemic helps researchers understand the intrinsic rules for distributing useful information and knowledge sharing in large-scale networks, and solve problems such as: How to spread valuable information in the network as wide as possible with minimum network flows (the reverse problem of the epidemic control)? How to enable a node to accurately receive the required information (for example, knowledge on real-world epidemic protection) in the network as quickly as possible with appropriate network flows? The social- and nature-aware characteristics help deepen our understanding of these issues.

Different types of flow obey different rules in the future interconnection environment. Organisms in the environment (called soft devices) process and spread information to form live scalable information flow networks [10]. They can represent software and digital devices, behave like both human and machine, and interact with each other to constitute a virtual society. Information flows—harmful or useful—form structures in the interconnection environment and hold obvious or hidden functions. Being aware or making use of this structure helps develop advanced functions to support effective information sharing and cooperation among information, knowledge, and service flows.

We live in a complex and diverse world. Computers initially used algorithms to create simplified models that generalized the real world. The social- and nature-aware computing environment should not step back to a complex world. Its main purpose is to find the new model that can better reflect the changing world—the fusion of nature, society, and the digital world, where rules will be different and evolving.

CONCLUSION

Just as the notion of Grid computing comes from the power grid, nature and society often reward scientists for inspiration. The practice of simulating the SARS epidemic provides us the ability to experience the overall evolution processes of a live scalable network. The degeneration phenomenon in its evo-

lution implicates the real evolution rules of versatile self-organization networks. The e-science Knowledge Grid environment supports advanced services for effective scientific activities based on effective information, knowledge, and service sharing.

The rules of an epidemic and e-epidemic further inspire us to explore the fusion and evolution of culture within semantic-rich social networks, from ancient to present, from real artifacts into the digital virtual world. **C**

REFERENCES
1. Chen, T.M. and Robert, J-M. Worm epidemics in high-speed networks. *IEEE Computer 37*, 6 (June 2004), 48–53.
2. Balakrishnan, H. et al. Looking up data in P2P systems. *Commun. ACM 46*, 2 (2003), 43–48.
3. Ebel, H., Mielsch, L.-I., and Bornholdt, S. Scale-free topology of e-mail networks. *Physical Rev. 66*, (2002).
4. Forrest, S., Hofmeyr, S.A., and Somayaji, A. Computer immunology. *Commun. ACM 40*, 10 (Oct. 1997), 88–96.
5. Foster, I., Kesselman, C., Nick, J., and Tuecke, S. Grid services for distributed system integration. *IEEE Computer 35*, 6 (June 2002), 37–46.
6. Kleinberg, J. Navigation in a small world. *Nature 406* (2000), 845.
7. Kleinberg, J. and Lawrence, S. The structure of the Web. *Science 294*, 30 (2001), 1849–1850.
8. Lipsitch, M. et al. Transmission dynamics and control of Severe Acute Respiratory Syndrome. *Science 300*, (June 2003), 1966–1970.
9. Oscar, P. and Roychowdbury, V.P. Leveraging social networks to fight spam. *IEEE Computer 38*, 4 (Apr. 2005), 61–68.
10. Zhuge, H. and Shi, X. Toward the Eco-Grid: A harmoniously evolved interconnection environment. *Commun. ACM 47*, 9 (Sept. 2004), 79–83.
11. Zhuge, H. *The Knowledge Grid*. World Scientific, Singapore, 2004.
12. Zhuge, H. The future interconnection environment. *IEEE Computer 38*, 4 (Apr. 2005), 27–33.

HAI ZHUGE (zhuge@ict.ac.cn) is a professor and director of the Key Lab of Intelligent Information Processing at the Institute of Computing Technology in Chinese Academy of Sciences, Beijing, China. He is the chief scientist of the China's National Basic Research Program's Semantic Grid Project and the founder of the China Knowledge Grid Research Group (kg.ict.ac.cn).