

An Interaction Based Composable Architecture for Building Scalable Models of Large Social, Biological, Information and Technical Systems

Introduction

This article describes our ongoing efforts to develop a global modeling, information and decision support cyberinfrastructure (CI) that will provide scientists and engineers novel ways to study large complex socio-technical systems. It consists of the following components:

- i. High-resolution scalable models of complex socio-technical systems;
- ii. Service-oriented architecture and delivery mechanism for facilitating the use of these models by domain experts;
- iii. Distributed coordinating architecture for information fusion, model execution and data processing; and
- iv. Scalable data management architecture and system to support model execution and analytics
- v. Scalable methods for visual and data analytics to support analysts.

To guide the initial development of our tools, we are concentrating on agent-based models of inter-dependent societal infrastructures, spanning large urban regions. Examples of such systems include: regional transportation systems; regional electric power markets and grids; the Internet; ad-hoc telecommunication, communication and computing systems; and public health services. Such systems can be viewed as organizations of organizations. Indeed, functioning societal infrastructure systems consist of several interacting public and private organizations working in concert to provide the necessary services to individuals and society. Issues related to privacy of individuals, confidentiality of data, data integrity and security all arise while developing microscopic models for such systems. See [1, 2, 3] for additional discussion (also see Figure 1).

Karla Atkins
 Christopher L. Barrett
 Richard Beckman
 Keith Bisset
 Jiangzhou Chen
 Stephen Eubank
 Annette Feng
 Xizhou Feng
 Steven D. Harris
 Bryan Lewis
 V.S. Anil Kumar
 Madhav V. Marathe
 Achla Marathe
 Henning Mortveit
 Paula Stretz
 Virginia Polytechnic Institute and State University

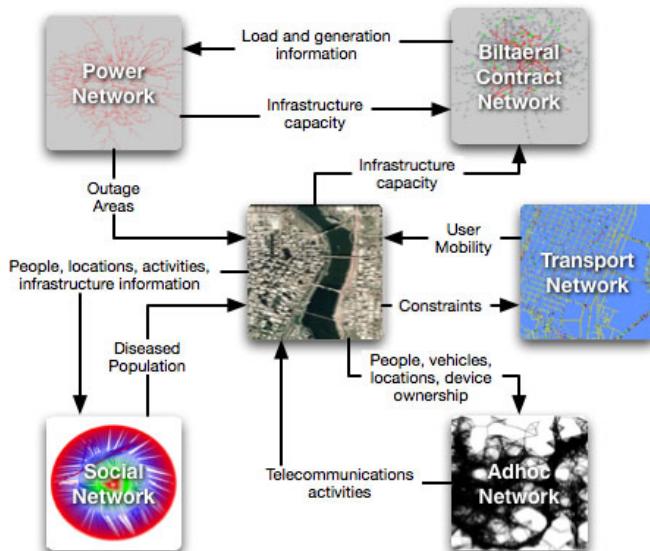


Figure 1. Schematic of societal infrastructure systems (adapted from [2]).

¹ Albert, R., Barabasi, A. "Statistical mechanics of complex networks," *Rev. Mod. Phys.* 74, pp. 47-97, (2002).

² Barrett, C., Eubank, S., Anil Kumar, V., Marathe, M. "Understanding Large Scale Social and Infrastructure Networks: A Simulation Based Approach," *SIAM news*, March 2004, Appears as part of Math Awareness Month on The Mathematics of Networks.

³ "Capturing Complexity Through Agent-Based Modelling," Special Issue of the *Proceedings of the National Academy of Sciences*, Vol. 99 (suppl. 3), 2002.

An Interaction Based Composable Architecture for Building Scalable Models of Large Social, Biological, Information and Technical Systems

The need to represent functioning population centers during complex incidents such as natural disasters and human initiated events poses a very difficult scientific and technical challenge that calls for new state-of-the-art technology. The system must be able to handle complex co-evolving networks with over 300 million agents (individuals), each with individual itineraries and movements, millions of activity locations, thousands of activity types, and hundreds of communities, each with local interdependent critical infrastructures. The system must be able to focus attention on demand and must support the needs of decision makers at various levels. The system must also support related functions such as policy analysis, planning, course-of-action analysis, incident management, and training in a variety of domains (e.g., urban evacuation management, epidemiological event management, bio-monitoring, population risk exposure estimation, logistical planning and management of isolated populations, site evacuations, interdependent infrastructure failures).

Constructing large socio-technical simulations is challenging and novel, since, unlike physical systems, socio-technical systems are affected not only by physical laws but also by human behavior, regulatory agencies, courts, government agencies and private enterprises. The urban transportation system is a canonical example of such interaction; traffic rules in distant parts of the city can have an important bearing on the traffic congestion in downtown, and seemingly “reasonable” strategies such as adding a new road somewhere might worsen the traffic. The complicated inter-dependencies within and among various socio-technical systems, and the need to develop new tools, are highlighted by the failure of the electric grid in the northeastern U.S in 2003. The massive power outage left people in the dark along a 3,700 mile stretch through portions of Michigan, Ohio, Pennsylvania, New Jersey, New York, Connecticut, Vermont and Canada. Failure of the grid led to cascading effects that slowed down Internet traffic, closed down financial institutions and disrupted the transportation; the New York subway system came to a halt, stranding more than 400,000 passengers in tunnels.⁴

The CI we are building was motivated by the considerations to understand the complex inter-dependencies between infrastructures and the society as described above. Over the past 15 years and in conjunction with our collaborators, we have established a program for modeling, simulation and associated decision support tools for understanding large socio-technical systems. The extremely detailed, multi-scale computer simulations allow users to interact among themselves as well as interact with the environment and the networked infrastructure. The simulations are based on our theoretical program in discrete dynamical systems, complex networks, AI and design and analysis of algorithms (see [2,5,6,7,8] and the references therein).

Until 2003, much of our efforts were concentrated on building computational models of individual infrastructures, see [2]. Over the last 7-10 years, significant advances have been made in developing computational techniques and tools that have the potential of transforming how these models are delivered to and used by the end users.^{9 10 11} This includes, web services, grid computing and methods to process large amounts of data. With the goal of harnessing this technology, since 2005, we have expanded the scope of our effort. In addition to building scalable models, we have also begun the development of an integrated CI for studying such inter-dependent, socio-technical systems. It consists of mechanisms to deliver the access to these models to end users over the web, development of a data management environment to support the analysis and data, and a visual analytics environment to support decision-making and consequence analysis (see [2,6]). The CI will provide social scientists unprecedented Internet-based access to data and models pertaining to large social organizations. In addition, the associated modeling tools will generate new kinds of synthetic data sets that cannot be created in any other way (e.g., direct measurement). The data generated by these methods will protect the privacy of individuals as well as the confidentiality of

⁴ US-Canada Power System Outage Task Force. Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations April 2004. <http://www.ksg.harvard.edu/hepg/Blackout.htm>

⁵ Barrett, C. L., Bisset, K., Eubank, S., Kumar, V. S. A., Marathe, M. V., Mortveit, H. S. “Modeling and Simulation of Large Biological, Information and Socio-Technical Systems: An Interaction-Based Approach,” *Proc. Short Course on Modeling and Simulation of Biological Networks, AMS Lecture Notes, Series: PSAPM*, 2007.

⁶ Barrett, C., Eubank, S., Marathe, M. “Modeling and Simulation of Large Biological, Information and Socio-Technical Systems: An Interaction Based Approach,” in *Interactive Computation: The New Paradigm*, D. Goldin, S. Smolka and P. Wegner Eds. Springer Verlag, (2005).

⁷ Eubank, S., Guclu, H., Anil Kumar, V. S., Marathe, M., Srinivasan, A., Toroczkai, Z., Wang, N. “Modeling Disease Outbreaks in Realistic Urban Social Networks,” *Nature*, 429, pp. 180-184 May (2004).

⁸ Barrett, C., Mortveit, H., Reidys, C. “Elements of a Theory of Computer Simulation: III,” *Applied Mathematics and Computation*, pp. 325–340, (2001).

⁹ Fox, G. “Grids Challenged by a Web 2.0 and Multicore Sandwich,” *CCGrid 2007 Keynote Talk* Rio de Janeiro Brazil May 15 2007.

¹⁰ Hayden, L., Fox, G., Gogineni, P. “Cyberinfrastructure for Remote Sensing of Ice Sheets,” *Proceedings of TeraGrid 2007 Conference*, Madison Wisconsin June 4-8 2007.

¹¹ *Grid Computing: Making the Global Infrastructure a Reality* edited by Fran Berman, Geoffrey Fox and Tony Hey. Wiley Publishers, March 2003.

An Interaction Based Composable Architecture for Building Scalable Models of Large Social, Biological, Information and Technical Systems

data obtained from proprietary datasets. This will enable social scientists to investigate entirely new research questions about functioning societal infrastructures and the individuals interacting with these systems. Everything, from the scope and precision of socio-technical analysis to the concept of collaboration and information integration, will change, as a dispersed framework that supports detailed interdependent interaction of very large numbers of complex individual entities that come into use and evolve. The tools will also allow policy makers, planners, and emergency responders unprecedented opportunities for coordination of and integration with the information for situation assessment and consequence management. This is important for planning and responding in the event of a large-scale disruption of the societal infrastructures.

Challenges

Current-generation High Performance Computing (HPC) based modeling environments for socio-technical systems are complex, large-scale special-purpose systems. These systems are readily accessible only to a limited number of highly specialized technical personnel in narrowly defined software applications. In other words, currently, web oriented distributed information CI for representing and analyzing large socio-technical systems simply do not exist. The vision of the CI under development is to change the model of delivery of HPC-oriented models and analytical tools to analysts. Just as the advent of search engines (e.g., Google) radically altered research and analysis of technical subjects across the board, the goal of the CI is to make HPC resources seamless, invisible, and indispensable in routine analytical efforts by demonstrating that HPC resources should be organized as an evolving commodity, and made accessible in a fashion as ubiquitous as Google's home page. Recently, grid based global cyber-infrastructures for modeling physical systems have been deployed. This is only a first step in developing what we would call *semantic complex system modeling*; we hope to make similar progress in the context of socio-technical systems. Developing such an infrastructure poses unique challenges that are often different than the ones faced by researchers developing the CI for physical systems. We outline some of these below.

1. Scalability: *The CI must be globally scalable.* The scalability comes in three forms: (i) allowing multiple concurrent users, (ii) processing huge quantities of distributed data and (iii) ability to execute large national-scale models. For example, simulating dynamical processes usually occurs on unstructured, time varying networks. Unlike many physical science applications, the unstructured network is crucial in computing realistic estimates, e.g., disease dynamics in an urban region. The unstructured network represents the underlying coupled, complex social contact and infrastructure. We need to simulate large portions of the continental United States – this implies a time varying dynamic social network of over 250 million nodes.

2. Coordination: *The CI should allow computational steering of experiments.* The systems needed by stakeholders are geographically distributed, controlled by multiple independent, sometimes competing, organizations and are occasionally dynamically assembled for a short period of time. Currently, web-enabled simulations and grid computing infrastructures for physical simulations have concentrated on massively parallel applications and loose forms of code coupling wherein large-scale experiments are submitted as batch jobs. Computational steering of simulations based on analysis is usually not feasible due to the latencies involved. Several socio-technical systems of interest can be formally modeled as partially observed Markov decision processes (POMDP) and large n-way games; a key component of POMDP and such games is that actions taken by an observer change the system dynamics (e.g., isolating critical workers during an epidemic). In other words, the underlying complex network, individual behavior and dynamics of particular processes over the network (e.g., epidemic)

An Interaction Based Composable Architecture for Building Scalable Models of Large Social, Biological, Information and Technical Systems

co-evolve. Such mechanisms require computational steering, which in turn requires CI for coordinating resource discovery of computing and data assets; AI-based techniques for translating user level request to efficient workflows; re-using data sets whenever possible and spawning computer models with required initial parameters; and coordination of resources among various users. *Computational steering* also occurs at a coarser level as well — in this case due to extremely large design space, we need to consider adaptive experimental designs. This is by and large infeasible on today's grid computing environments.

3. Data & Information Processing: *The CI should facilitate efficient data and information fusion and analysis.* The Internet has enabled the sharing of data in a simple and cost effective way, from the producers' side. Consumers of the data must still locate the appropriate data and deal with multiple incompatible data formats. The heterogeneity, volume and geographic distribution of data implies that social scientists, without the proper tools and use of database techniques, will be left to write custom programs that will tend to be less efficient than well crafted database and middleware methods. Unlike simulations of physical systems, models of socio-technical systems are usually data-intensive. Moreover, the data sets are being continually collected, refined, integrated and aligned to support ongoing analysis. Analogous to physical simulations, the output data is large and processing it is a computational challenge. More importantly, a POMDP model of socio-technical systems implies that a lot of data mining and analysis has to be done in concert with the simulation. This implies stringent computational requirements.

4. User Support: Development of appropriate analysis frameworks for users are needed, including user interfaces, high level formalisms to set up experiments, and visual and data analytics, which include methods for integrating heterogeneous databases to support multi-view visualization (e.g., disease spread in a geographic region and epidemic curves); methods for visualizing and analyzing large co-evolving coupled networks; and data mining and knowledge discovery tools to support analytical processes.

In addition, we need to develop environments and tools for simulation assisted decision support and consequence analysis. This includes methods for presenting results of analysis and simulations so as to avoid confirmation bias and framing effects, simulation based micro-economic analysis of decisions, and methods in risk analysis for ranking assets and understanding the inherent uncertainties in modeling such systems.

Overall Architecture & Current Status

Figure 2 shows a conceptual architecture of the overall system that we are developing. Simfrasructure assumes the role of coordination between all the constituent components. This includes high resolution models for simulating large socio-technical systems, SimDM: a distributed data management environment, the underlying data and compute grids that provide low level data and compute services. Simfrasructure uses (tuple/java)-spaces to achieve the desired coordination goals. Currently, we have operating models for public health, commodity markets, transportation, integrated telecommunication networks, urban populations and built infrastructure. These models can all run on high performance computing platforms. We are currently extending them to work on grid-like architectures developed as a part of the NSF funded Teragrid initiative.

An Interaction Based Composable Architecture for Building Scalable Models of Large Social, Biological, Information and Technical Systems

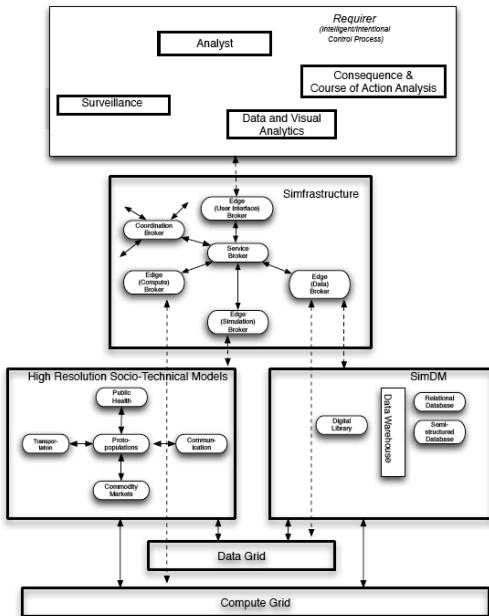


Figure 2. Overall cyber-infrastructure architecture for simulation based decision support and consequence analysis of large societal infrastructures.

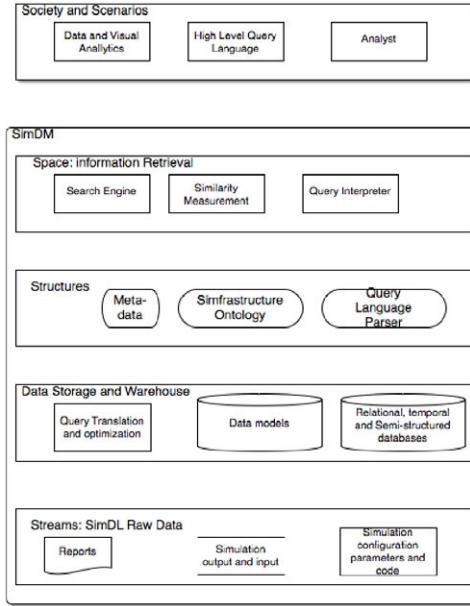


Figure 3. Schematic architecture of SimDM: data management module based on the 5S specification.

Current Infrastructure Models: Currently, modules for public health, telecommunications, commodity markets and urban social networks have been developed and integrated within the Simfrastructure framework [5, 6, 12, 13, 14]. Simdemics is a module that supports public health epidemiology. Simdemics itself has three different implementations of methods that simulate the spread of infectious diseases. Sigma is a highly scalable, web-based, service-oriented modeling framework for analyzing large generic markets for commodities such as electricity, oil, corn, as well as for allocating distributed computer resources in a utility data center. It supports large-scale synthetic and human economic experiments and studies related to bargaining, learning, cooperation, and social and risk preferences. It can also be used as a tool to study appropriate designs for marketing bandwidth in an unlicensed radio spectrum. The integrated tele-communication modeling environment consists of analytical and simulation-based modeling tools for design and analysis of next-generation computing and communication systems that are based on packet switched network technology. Examples of such systems include mesh networks deployed in urban and rural communities, vehicular ad hoc networks, hybrid, cellular, mesh and sensor networks. Finally, the module for generating urban social contact networks generates high fidelity synthetic networks consisting of people, locations and their interactions. The kinds of interactions determine the specific social network that is created. In addition to these, TRANSIMS an urban transport module developed by our group and team members at Los Alamos National Laboratory, is also integrated within the framework. A key feature of the overall architecture is its ability to easily integrate other infrastructure models within the framework. For instance, the Urban infrastructure suite (UIS) developed at Los Alamos can be integrated within the current framework.

SimDM is an integrated data management environment. It follows the 5S (Streams, Structures, Spaces, Scenarios, and Societies) framework that defines the meta-model for a minimal digital library. A conceptual architecture is shown in Figure 3 above. See [15, 16] for additional discussion.

¹² Atkins, K., Barrett, C., Beckman, R., Bisset, K., Chen, J., Eubank, S., Anil Kumar, V. S., Lewis, B., Macauley, M., Marathe, A., Marathe, M., Mortveit, H., Stretz, P. "Simulated Pandemic Influenza Outbreaks in Chicago," *NIH DHHS Study Final report*, 2006.

¹³ Barrett, C., Smith, J. P., Eubank, S. "Modern Epidemiology Modeling," *Scientific American*, March (2005).

¹⁴ Barrett, C., Beckman, R., Berkbigler, K., Bisset, K., Bush, B., Campbell, K., Eubank, S., Henson, K., Hurford, J., Kubicek, D., Marathe, M., Romero, P., Smith, J., Smith, L., Speckman, P., Stretz, P., Thayer, G., Eekhout, E., Williams, M. D. "TRANSIMS: Transportation Analysis Simulation System," *Technical Report LA-UR-00-1725*, Los Alamos National Laboratory Unclassified Report, 2001. An earlier version appears as a 7 part technical report series LA-UR- 99-1658 and LA-UR-99-2574 to LA-UR-99-2580.

¹⁵ Bailey-Kellogg, C., Ramakrishnan, N., Marathe, M. "Spatial data mining to support pandemic preparedness," *SIGKDD Explorations* 8: 80-82, 2006.

¹⁶ Barrett, C., Bisset, K., Eubank, S., Fox, E., Ma, Y., Marathe, M., Zhang, X. "A Scalable Data Management Tool to Support Epidemiological Modeling of Large Urban Regions," *European Conference on Digital Libraries (ECDL)*, pp. 546-548, 2007.

An Interaction Based Composable Architecture for Building Scalable Models of Large Social, Biological, Information and Technical Systems

It stores streams of textual bits from files or databases and audio/video sequences. Challenges arise from enforcing proper structures over heterogeneously structured digital objects with close conceptual relationships. In our prototype implementation, we used RDF-based metadata, which defines semantic contents of objects and relationships among them. The metadata constructs a knowledgebase for Simfrasrtucture, on which a browsing service could be based. Simfrasrtucture objects contain both textual information and real number parameters.

Simfrasrtucture serves the role of coordinating these diverse simulations, the data management tool, the visual and data analytic tools, and the end users. It is organized around the Software as a services (SAS) paradigm. Currently, it uses Javaspaces as the implementing construct, but the basic concepts are generic and readily implemented using other similar languages. We will use the terminology in Gelernter and Carriero^{17 18} for describing Simfrasrtucture. The *asynchronous ensembles* in our architecture consist of simulation models, databases, GUIs and analytical tools. The basic concept within Simfrasrtucture is that of *brokers*: *coordinating processes responsible for achieving a desired workflow by appropriately invoking appropriate asynchronous ensembles*. Brokers use *associative memory* for communicating data objects between them. In Javaspaces this is called a blackboard. For computational efficiency and security, these blackboards are generally distributed and organized hierarchically. Brokers are also organized hierarchically; this hierarchy captures calling rules. As in tuple-spaces or Javaspaces, brokers place appropriate data objects in the associative memory. Our architecture uses the generative communication paradigm; brokers act as coordinators for this purpose. Brokers are responsible for understanding what information needs to be communicated between various asynchronous ensembles. They are lightweight processes that are assigned the task of requesting information from various ensembles, communicate information/data between ensembles by using blackboard and in the end achieve a given workflow. Important parameters are that of computational efficiency, memory requirement and accuracy. In our envisioned architecture, achieving a given functionality has to account for these parameters; brokers call appropriate computation and evaluation processes to conclude if the data object returned conforms to the required specification. We have chosen to use a tuple-space model in contrast to a message passing model for our coordinating system for the following reasons:

- Brokers in general will not know how a specific request can be satisfied, therefore, it uses common associative memory as a way to broadcast its request. Brokers that can invoke appropriate processes to fulfill these requests using **in** or **rd** like primitives available in Linda tuple spaces. This was one of the central features of tuple-space like constructs and is very useful in our setting.
- The broker requests are highly asynchronous; in general, requests are generated on demand when a specific analysis needs to be done by an analyst. At that time, we have very little control over the specific computing and data resources at our disposal.
- Broker based architecture allows us to develop solutions that protect participating institutions' IP and security requirements. Since all communication happens among brokers and not directly between services, organizations need not have knowledge either of the entire problem or all of the resources being used to solve the problem. By using a trusted third-party to host the computation, one organization may provide a proprietary model that uses proprietary data from a second party, without either organization needing a trust relationship with the other.

¹⁷ Gelernter, D., Carriero, N. "Coordination Languages and their significance," *CACM*, 1992, 35(2).

¹⁸ Gelernter, D., Carriero, N. "Linda in Context," *CACM*, 1989, 32(4).

An Interaction Based Composable Architecture for Building Scalable Models of Large Social, Biological, Information and Technical Systems

In the current implementation of Simfrastructure, we have three kinds of brokers:

- *Edge brokers*: mediate access to a particular resource (simulation, data, service, etc), removing the need for the resource to communicate directly with any other resource
- *Service brokers*: coordinate the edge brokers in response to satisfy a given query
- *Coordination brokers*: coordinate the overall workflow

There are several kinds of edge brokers: simulation edge brokers meant for interacting with simulations; data management edge brokers meant to access the data management module; the visual and data analytics brokers; surveillance data access brokers, etc. These edge brokers can in turn call data and compute grid primitives as necessary. A formal grammar specifies the call structure and the various access rules. Simfrastructure in concert with the simulation models and SimDM are designed to support computational steering; this is crucial for the class of applications we are interested in studying.

Illustrative Scenarios

It is valuable to go through an illustrative scenario of how the CI might be used in practice by analysts and policy makers for planning and response to natural or human-initiated crises. The CI been used to support several user defined studies over the past 10 years; see [3, 7, 19]. The scenario illustrates the need to address each of the challenges discussed above. It also highlights the need to solve the complete problem rather than solving it piecemeal. Solutions need to be practical and usable.

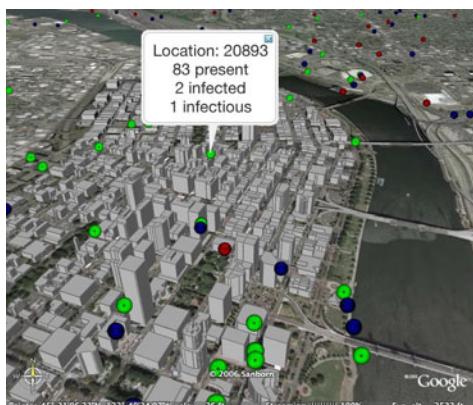


Figure 4. A sample graphic available to the analyst when using the cyber-infrastructure in the illustrative scenarios. Note the spatial detail that can be generated.

¹⁹ Atkins, K., Barrett, C., Beckman, R., Bisset, K., Chen, J., Eubank, S., Anil Kumar, V. S., Lewis, B., Marathe, A., Marathe, M., Mortveit, H., Stretz, P. "An analysis of layered public health interventions at Ft. Lewis and Ft. Hood during a pandemic influenza event," TR-NDSSL-07-019, 2007.

Situational Awareness and Consequence Analysis in the Event of Epidemics

In this scenario, during a heat wave in a city, an adversary shuts down portions of the public transit system and a hospital emergency room during the morning rush hour. They spread a harmless but noticeable aerosol at two commuter rail stations. These events, occurring nearly simultaneously, foster a chaotic, if not panic-stricken, mood in the general public. Disinformation released via the mass media enhances the perception of an attack. Simulations of epidemics on social contact networks combined with simulations of urban population mobility and other infrastructure simulations can be used for situation assessment and course of action analysis. Simfrastructure first calls the data broker to see if the network is already available. Assuming the answer is no, Simfrastructure then calls the population mobility broker for constructing the dynamic social network. This in turn will involve fusing information about census data, data on individual activity, and location data from commercial and public databases.

An Interaction Based Composable Architecture for Building Scalable Models of Large Social, Biological, Information and Technical Systems

These are the people who would show up first for treatment if indeed a chemical or biological attack had occurred. They also would serve as the subpopulation to seed our epidemiological simulations. Simfrasstructure calls data mining tools and Simdemics to achieve these tasks.

Biases in their demographics compared to a random sample of the population will induce persistent biases in the set of people infected at any time. We estimated the demand at hospitals, assuming that people would arrive at a hospital near their home or current location. We also estimated the demographics of casualties under an alternative scenario during only a heat wave. Historically, the most likely casualties of a heat wave are elderly people living alone with few activities outside the home.

This information, combined with demographic and household structure data, allowed us to estimate demand for health services created by the heat wave by demographic and location. For situation assessment, we noted the obvious differences between these two demand patterns. In an actual event, comparison with admissions surveillance data would allow quick disambiguation between the two situations. We estimated the likely spread of disease for several different pathogens by demographic and location. Furthermore, we implemented several suggested mitigating responses, such as closing schools and/or workplaces, or quarantining households with symptomatic people. Knowledge of the household structure permits an exceptionally realistic representation of the consequences of these actions. For example, if schools are closed, a care-giver will also need to stay home in many households.

Conclusions and Summary

We described our work in progress that aims to build a scalable CI to study large socio-technical networked systems. The goal of the CI is to provide seamless access to HPC-based modeling and analysis capability for routine analytical efforts. It consists of (i) high-resolution models, tools for decision making, and consequence analysis, (ii) service-oriented architecture and delivery mechanism for facilitating the use of these models by domain experts, (iii) distributed coordinating architecture for information fusion, model execution and data processing, and (iv) scalable methods for visual and data analytics to support analysts.

Due to space considerations, we have not discussed peta-scale computing and data grids that will serve as the underlying technology. Much remains to be done to develop the CI. Researchers across the world are developing new tools in web services, tools and CI for various problem domains.^{9 10 11} We hope to build on these advances. 

Acknowledgements

The work is partially supported by an NSF HSD grant, the NIH MIDAS project, the CDC center of excellence, DoD and a Virginia Tech internal grant. We thank our current and past collaborators, on related topics, especially, Douglas Roberts, Aravind Srinivasan, Geoffrey Fox, Arun Phadke and Jim Thorp, students in the Network Dynamics and Simulation Science Laboratory and members of the TRANSIMS and NISAC projects at Los Alamos National Laboratory. Finally, we thank Suman Nadella and Pete Beckman for inviting us to submit an article for this *CTWatch Quarterly* issue.