Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

DYNAMIC DATA DRIVEN APPLICATION SYSTEMS: RESEARCH CHALLENGES AND OPPORTUNITIES

Richard Fujimoto

School of Computational Science
and Engineering
Georgia Institute of Technology
Atlanta, GA 30332, USA

Joseph Barjis

Department of Industrial
and Systems Engineering
San Jose State University
San Jose, CA 95192, USA

Erik Blasch

Air Force of Scientific Research
Air Force Research Laboratory
Arlington, VA, 22203, USA

Wentong Cai

School of Computer Science and Engineering
Nanyang Technological University
Singapore 639798

Dong Jin

Department of Computer Science
Illinois Institute of Technology
Chicago, IL 60614, USA

Seunghan Lee
Young-Jun Son

Systems and Industrial Engineering
University of Arizona
Tucson, AZ 85721, USA

ABSTRACT

Dynamic Data Driven Applications Systems (DDDAS) is a paradigm where data is dynamically integrated into an executing application, and in reverse, the application dynamically steers the measurement process in a feedback control loop. Since its inception in 2000, the DDDAS concept has been successfully applied to a host of application areas. New technologies are emerging such as big data, the Internet of Things, and cloud/edge computing. With these trends DDDAS is poised to have large-scale impacts in areas such as smart cities, manufacturing, health care, and security, to name a few. Each author describes their views concerning the important research challenges facing the DDDAS paradigm and opportunities for impact in the years ahead.

1 INTRODUCTION

Dynamic Data Driven Applications Systems (DDDAS) was proposed by Frederica Darema as a paradigm where measurement data from an operational system is dynamically incorporated into an executing model of that system, and computational results from the model are then used to guide the measurement process (Darema 2004). Measurement data is typically used by the model to improve its accuracy and/or execution time, while computational results produced by the model help steer the measurement process itself. The DDDAS paradigm has been applied to applications such as emergency response, transportation, electric power grids, manufacturing, surveillance, and many others. Information and references to recent work in DDDAS are summarized in (DDDAS 2018). The following sections present position statements by several leading researchers concerning key DDDAS research challenges.

The sections that follow span a variety of issues ranging from design methodologies, system design, modeling approaches, and applications from several different perspectives. Blasch begins by first presenting the DDDAS cycle of sensing, learning, and adapting and outlines key underlying technologies. He provides a broad discussion of DDDAS research challenges highlighting the need for advances in theory, algorithms and computation. He discusses challenges in data science, autonomy and smart sensing. Fujimoto focuses on the development of DDDAS for mobile and high performance computing systems. He highlights the need to address research issues related to power management and energy consumption. Cai emphasizes modeling issues. He discusses challenges in data-driven agent-based modeling and simulation, especially when such systems involve human participants. The remaining authors focus on DDDAS research challenges in the context of specific application domains. Barjis begins by discussing DDDAS in the context of agile system design. Specifically, he highlights the implications of incorporating data analytics in software and system architectures. Jin discusses challenges in creating secure control and communications systems for smart electric power systems. Lee and Son highlight challenges in the context of smart border security systems. Collectively, the authors identify many areas where additional research is required. This paper highlights the breadth of research challenges that remain, spanning many different aspects of DDDAS design, development, and deployment.

2 DDDAS AUTONOMY (ERIK BLASCH)

Building on the DDDAS use of high-dimensional modeling to augment the feedback loop for real time execution, the future areas of DDDAS include (1) data science learning, (2) autonomy through adaptation, and (3) systems design with smart sensing – as shown in Figure 1 (Blasch et al. 2018). The developments of the integration of the instrumentation, models, and software to enable the development of DDDAS include: ***theory, algorithms, and computation***. The theory includes mathematical advances (e.g., retrospective cost modeling); while the algorithms support new methods (e.g., ensemble Kalman filter, Particle filter, optimization techniques) (Dunik et al. 2015). The computational considerations align with the developments in the continuing networked society such as non-convex optimization, data flow architectures, and systems design (Blasch et al. 2012; Imai et al. 2017; Li et al. 2017) to support modeling and simulation (Mustafee et al. 2017).

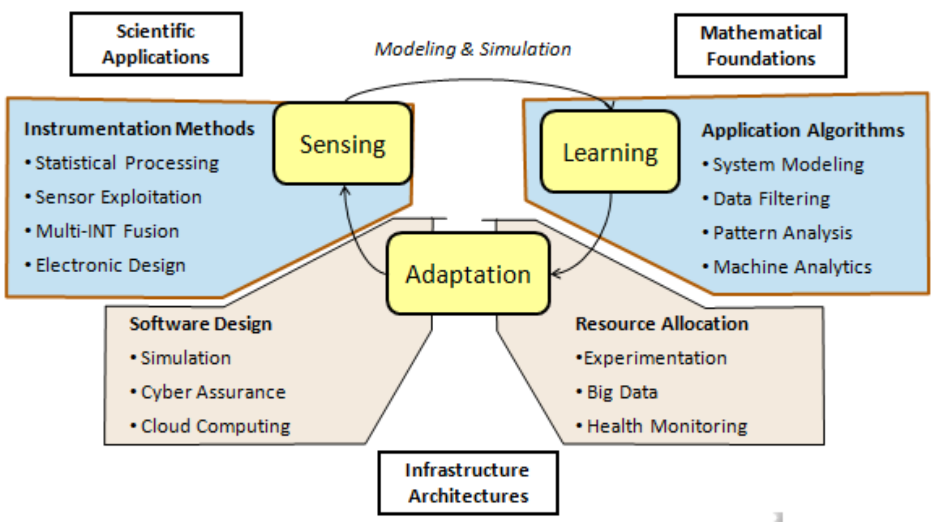

Figure 1: Dynamic Data Driven Application Systems (DDDAS) Future Areas.

Data movement and ***data science*** are future efforts aligned with the growth of artificial intelligence (AI), machine learning, and deep networks. These growing areas of interest follow from the recent trend in big data. The original DDDAS paradigm calls out *big data* as an emerging theme that utilizes algorithms, models, and computation to harness data availability. Algorithms still need to be adapted from the static to the dynamic environment. For example, neural networks do interpolation through modeling, but are not good at extrapolation to changing environments nor explanation of resolved decisions. If, on the other hand, DDDAS instrumentation methods capturing data are integrated with high-dimensional modeling of the situation context, then such constructs within AI can be realizable for design efforts.

Autonomy includes building on the data at rest, data in motion, and data in use concepts. While the data concepts (i.e., at rest, in motion, and in use) were promoted by the software community, these labels were mostly for the movement of data and not the processing of the data. The processing of the data, augmented with modeling can be a hallmark of future DDDAS autonomous systems. Autonomy at Rest (AAR) leverages data science to combine or fuse the data; Automomy in Use (AIU) includes the data being processing in real time; while Autonomy in Motion (AIM) supports the interaction among platforms such as unmanned aerial vehicles (UAVs). Autonomous systems dynamically interact with the environment, so there is a need for not only complex modeling, but also methods in which real-time distributed sensing updates the models. Together, modeling, sensing, and data movement are future trends in DDDAS to achieve autonomous solutions.

The third growth area is in ***smart sensing*** through networked systems and software, or architectures to move and process data with high performance computing over a wide variety of sensors. The coordination of social modeling, internet of things (IoT), cyber physical systems, and power grids require systems and software developments to coordinate the dynamic data. The DDDAS efforts will expand from DDDAS principles, while leveraging the developments in autonomy and learning. Additionally, the sensing includes not only the data exploitation, but also information collection.

The *challenges DDDAS* seeks to advance include data modeling, context processing, and content application (Snidaro et al. 2016; Zheng et al. 2018). To bring together data, context and content requires addressing issues in *model fidelity*, dimension, and usability such as how many parameters are needed for system control. When data is collected, it needs to be preprocessed to determine whether its inherent information *matches* the context. One example includes clutter reduction, sensor registration, and confuser analysis in vehicle tracking. Finally, another key challenge is that of sampling, as shown in Figure 2. *Sampling* is the multiresolution needed to monitor the situation, environment, and network context to explain the content desired.

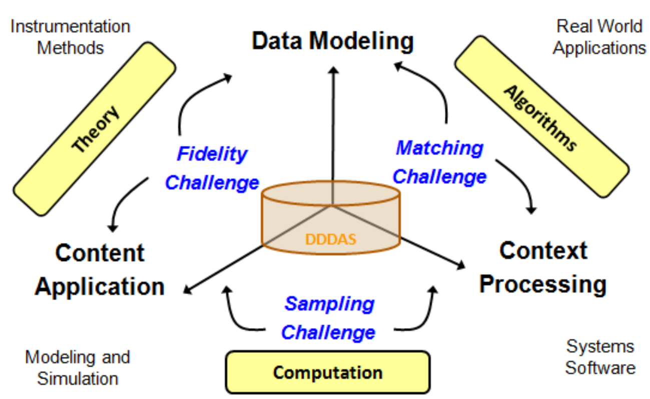

Figure 2: Dynamic Data-Driven Application Systems (DDDAS) Challenges and Processes.

DDDAS applications will push various technologies to leverage models and big data. Advances will include: (1) *theory* (e.g., data assimilation, process modeling and filtering, and estimation); (2) *methods* (e.g., structural analysis for structural health monitoring, systems control for component processing, and image computing for situation evaluation); and (3) *design* (situation awareness through environmental assessment, energy awareness such as power grids, and cyber awareness concerning privacy and security protections) (Blasch et al. 2018).

3 ENERGY MANAGEMENT IN MOBILE DDDAS (RICHARD FUJIMOTO)

Three important trends will have significant impacts on future DDDAS deployments. First, the explosion of communicating computational devices embedded in the physical world, sometimes referred to as the Internet of Things (IoT), will continue into the foreseeable future. This widespread deployment of computing capability enables enhanced exploitation of the DDDAS paradigm. Second, the computational capabilities of embedded computing systems will continue to grow. One example is the *micro-cluster*, a high performance computing cluster composed of processors commonly found in mobile computing devices such as smart phones. The micro-cluster has the potential to have a transformational impact in embedded computing analogous to the creation of Beowulf clusters and the advent of cluster computing in the 1990's stemming from the widespread deployment of high performance workstations. Third, increased deployment of mobile platforms such as small commercial drones and other autonomous vehicles creates new physical platforms in which to exploit the DDDAS paradigm. These trends suggest that mobile DDDAS will be increasingly important in the years ahead.

These trends suggest that DDDAS deployments will increasingly utilize computational models that are embedded in the physical world itself rather than being relegated to back-end compute servers. This approach offers several advantages over back-end or cloud-based deployments. For example, the computational models can operate in close proximity to data sources, allowing computations to utilize disaggregated data. Spatial proximity to the physical system allows the realization of more tightly coupled DDDAS control loops requiring low latencies. The distributed approach reduces reliance on wide area networks offering greater resilience to network outages and cyber-attacks. Further, local processing of data can help address privacy concerns compared to approaches requiring sensitive information to be transported and stored within the cloud.

As DDDAS migrates to mobile computing devices the energy consumed by computational models and communications becomes an important concern. Reductions in *energy consumption* can lead to increased battery life, and/or enable the use of smaller, lighter batteries. Further, in high performance computing (HPC) applications such as those executing on micro-clusters heat generation may be a concern, necessitating caps on power consumption. Indeed, power consumption is an area of increased concern in data centers and is gaining in importance in the HPC community. While there is a substantial literature in energy- and power-aware computing, prior work to date has largely focused on the lower levels of the computing stack considering hardware techniques such as dynamic voltage and frequency scaling (DVFS), operating systems issues such as scheduling, networking protocols, and compiler and language issues. Other work has focused on reducing energy consumption in specific applications. The amount of research aimed at understanding and reducing energy and power consumption in algorithms and computational models for DDDAS deployments is comparatively modest.

Effective deployment requires careful consideration of power and energy concerns in conjunction with execution time requirements. A goal in the design of energy efficient DDDAS might be to minimize energy consumption while meeting execution time requirements. It may be noted that in many cases minimizing energy consumption is consistent with minimizing execution time. Power consumption can be broken down into static consumption resulting from leakage currents where power is consumed regardless of what computations are being performed, and dynamic consumption that is a function of the circuits that must be utilized to complete the computation. As such, minimizing computation time will typically minimize energy consumption, though that is not always the case.

There are several aspects of a DDDAS deployment where energy consumption is poorly understood, highlighting areas where further research. These are:

- *Simulation models and software*. It has long been understood that the choice of simulation modeling approach can have a large impact on execution time. The most appropriate choice depends on the level of detail required to meet the needs of the simulation's predicative capability, balanced against the time and energy required to execute the model. For example, many different abstractions exist to model a transportation network. These are often characterized as macroscopic, mesoscopic, and microscopic based on the level of aggregation. One study (Neal et al. 2016) compared the energy consumed by queueing network and cellular automata based data driven traffic simulations. The choice of data structure used to implement the future event list in a discrete event simulation similarly has ramifications in terms of energy consumption. The amount of energy is impacted by the pattern of memory references which impacts the energy consumed in the memory hierarchy, and this will vary from one data structure to another (Neal and Fujimoto 2018).

- *Data distribution*. Data distribution mechanisms are needed to transmit measurement data to a data-driven simulation, to distribute data among the simulation processes if a parallel or distributed simulation approach is used, and to effect changes in the operational system as the result of model computations. For example, one approach for data distribution are the declaration management and data distribution management (DDM) services defined in the High Level Architecture standard (IEEE, 2010). In (Neal and Fujimoto 2017) the energy requirements for computation and communications to implement the HLA DDM services are examined where a tradeoff between these two elements was observed for different implementation approaches.
- *Synchronization*. In parallel and distributed discrete event simulations a synchronization algorithm is required to ensure the simulation computation produces the same results as the corresponding sequential implementation. Synchronization algorithms are generally classified into two categories termed conservative and optimistic. The energy consumption properties of synchronization algorithms are not well understood. In (Biswas and Fujimoto 2018) a class of synchronization algorithms terms *zero-energy synchronization* is defined where the amount of energy consumed by a distributed simulation is the same as that required by the simulation computation and that needed for event communication; i.e., no additional energy is required for the synchronization algorithm. A low energy version of the well-known YAWNS synchronization algorithm termed LEY (low energy YAWNS) is proposed. It is argued that in principle this approach can yield zero energy synchronization with certain constraints on the simulation application. Empirical data is presented showing that LEY utilizes much less energy than a version of YAWNS not optimized to reduce energy consumption, and is observed to yield energy consumption approaching that of a zero energy synchronization algorithm.

Finally, another area of research concerns the integration of mechanisms described above to create energy efficient runtime infrastructure (RTI) software for DDDAS. While low energy implementations of the various services required to implement such middleware are needed, these implementations must be integrated to create a software base to support mobile DDDAS.

4 DYNAMIC DATA-DRIVEN AGENT-BASED MODELLING AND SIMULATION (WENTONG CAI)

Although traditional simulation analysis can be used to generate and test possible scenarios, it suffers from a long cycle-time for model update, analysis and verification. It is thus difficult to carry out prompt what-if analysis to respond to abrupt changes in the physical systems being modelled. The DDDAS paradigm solves the what-if problem by having the simulation system and the physical system interact in a mutually beneficial manner. The simulation system benefits from real-time input data which are used to adapt the model and the physical system benefits from improved performance or better understanding obtained from the analysis of simulation results.

Under the DDDAS paradigm, data analytics can be used to exploit information to identify patterns, create possible what-if scenarios, make predictions about the future, and determine actions based on the accurately predicted outcomes. In this way, we can draw conclusions from both historical data and the predictive power of the simulation. The importance of data analytics techniques in modelling and simulation (M&S) and how they can be applied to various phases of M&S have been addressed in (Tolk 2015; Taylor et al. 2015; Nelson 2017). In this section, we specifically explore the synergy amongst DDDAS, data analytics, and *agent-based modelling and simulation* (ABMS) and discuss the challenges of applying dynamic data-driven techniques in various stages of ABMS.

Applications integrating analytics and simulation often require data to be acquired on the fly for such purposes as validating simulation results, creating simulation scenarios, and refining and building simulation models. In many cases, the data acquisition may not be feasible with traditional methods and human involvement is mandatory. For example, the human behavioral models would need to be built in many agent-based simulations in order to carry out analysis of complex adaptive systems (e.g., socio-ecological systems). However, without any implementation or occurrence of the hypothesized scenarios in reality, it would be impossible to collect human behavioral data directly from the physical world. Thus, one big challenge, particularly for agent-based modelling and simulation, is to design new data acquisition methods for constructing reliable human behavioral models to drive simulations.

A possible approach is to create *virtual environments* in the cyberspace to emulate the hypothesized scenarios and invite people to participate in the virtual environments through modern social media in order to acquire their behavioral data (Kretz et al. 2011; Viswananthan et al. 2017). In addition, applications like pandemic simulation require data acquisition at the scale and extent that are beyond the capabilities of dedicated data acquisition infrastructures. A potential approach is to harness human ingenuity by means of crowd-sourcing or participatory-sensing to assist in data acquisition. A significant challenge is to deal with the inherent uncertainty and errors in the acquired data.

One fundamental issue in agent-based modeling and simulation is to develop behavior models of agents and find their correct parameter settings, so that the simulation can match the behavior observed in the physical system. Traditional methods require in-depth domain knowledge and involve a great amount of manual effort. Using data analytics techniques, persistent patterns and transient patterns in the data can be discovered. Persistent patterns will become building blocks for the model, while transient patterns will become interactions between building blocks. Some preliminary work has been carried out using data-driven approach to create models for agent-based crowd simulation (Ma et al 2016; Zhong et al. 2017). However, how data analytics can be used in general to extract useful knowledge and insight from the data to facilitate model development for agent-based simulation still remains an open research question.

With the technology development in sensors, detectors, and mobile devices, live-streamed data become easily available. As such, it is possible to have a quantitative agreement between simulation and the physical system at very detailed level. To achieve this, methods need to be developed to align the simulation with the physical system so that data from the physical system can be used to keep the simulation adaptively close to what happens in the real-world. This is clearly required, for instance, for real-time agent-based traffic simulation since traffic situations vary throughout the day or change when serious traffic accidents occur. Without model recalibration or incremental model refinement based on real-time feedback derived from live-streamed data, the forecast from the simulation in this case would be useless. To this end, data assimilation techniques have been used to incorporate real-time data into a running simulation (Wang and Hu 2015) or to calibrate agent-based models (Ward et al. 2016) to generate more accurate results.

Finally, to study the effect of parameters that influence an agent's behavior or to investigate different what-if scenarios, an agent-based simulation study often involves a large number of simulation runs. To reduce resource utilization, it is more important to optimize the number of simulations in a simulation-based study than the performance of individual simulations (Fujimoto et al. 2017). With high fidelity and a microscope model, an agent-based simulation execution may generate large amounts of data during the simulation. This creates an opportunity for "just-in-time" analysis of the simulation data to infer the usefulness (or utility) of a simulation run. The simulation execution can be pre-emptied to reduce resource utilization if its utility is low. Data analytics techniques can also be used to manipulate or steer simulation experiments on the fly. For example, an approach has been described in (Feldkamp et al. 2017), which takes flow of simulation runs as steamed data and dynamically discovers correlation between input parameters and simulation outputs using streamed k-mean clustering and decision tree classification. As demonstrated in a use case scenario, the approach is able to make correct assumptions about the underlying system with just a fraction of the total number of simulation runs required in the experiment.

5 DDDAS APPLICATION TO AGILE USING TELEMETRY (JOSEPH BARJIS)

While the DDDAS paradigm is by now well studied, researched, and published; it is new application areas that require scientific explorations. Along with the service delivery, transportation, manufacturing, surveillance, and other areas, Agile System/Software development is a new research area, where the potential of the DDDAS merits profound research.

In the last two decades, nothing shaped and transformed the practice of software and system development as radically as the *Lean-Agile* practice. In the core of Agile practice are iterative development and continuous adjustment of the system scope and functionality based on the feedback of earlier delivered pieces of functionality (Leffingwell 2011). To this end, the notion of feedback loop constitutes the very essence of Agile practice.

On the other hand, in an information age that data becomes new natural resources (Picciano 2014) and both application execution and user interaction contribute to the pool of big data, automatic collection of data must be integral to the development of new applications. Data collection should be an embedded capability of each software feature on top of the business function that they support. So, they not only deliver business functionality but also collect data on their own performance. As natural resources, raw data present less value if they are not accessible, usable, and customized for user needs.

In addition to these two arguments, the importance of data analytics and predictive analytics is ever growing in new system development as well as enhancing existing systems.

In the Agile system and software development practice, new applications are developed by continuously adding new functionalities, usually referred to as business Features (see Figure 3, dark blue color items in the product backlog) and improving the application architecture and environment, usually referred to as *Enablers* (see Figure 3, dark red items). Next to that is continuous deployment, including in the production environment, which is accomplished when development and operation teams work for the same goal, fast value delivery to customers. This collaboration is coined as *DevOps*. This is where automatic data collection is important to gain insight into the behavior of each change introduced through a new Feature to the existing application, and to see whether the change impacts the performance of the system or application as a whole. Furthermore, these data help to see the business use of a Feature and whether the benefit hypothesis for that Feature holds true.

To weave data analytics capability right into the fabric of new applications, i.e. Agile practice, in addition to Features and Enablers; one should be encouraged to add data analytics items in the product backlog (Figure 3, see green color items). These *data analytics backlog* items will comprise building blocks that a prospective simulation model can be built upon. Such a simulation model is used to predict the application behavior partially based on real data injected from the application execution and partially from mock or archival data to fill in for the Features that are not yet developed or deployed.

In early phases of application development, simulation models can be used to represent not yet existing Features or system components. While the features deployed in production will show an actual behavior of the application so far developed, the *simulation model will animate and predict* the whole behavior of the application once it is fully developed and the remaining features are implemented. These applications can be fairly complex, expensive, multiyear development, and cross boundaries of many functions and departments.

In essence, here the simulation model plays a multi-purpose role, including filing in the gaps of yet incomplete features. Also, simulation outcomes contribute towards measurement, which is critical in Agile software/system development. This way, measurement and metrics supported and enhanced by the simulation results can be cost effective on one hand, and allowing to see the results of otherwise a very lengthily data collection process that takes days or longer.

From an Agile software and system development perspective, the most important use of the embedded simulation model is creation of a feedback loop (see Figure 3), which evolves as the application development is further and further advanced by adding new features. This novel application of simulation based on real data and on top of work-in-progress may suggest where to improve performance, which future features are in most demand, and whether users prefer certain ways over others.

Furthermore, the intentional introduction of chaos to study robustness and resilience of the application can be first studied at the simulation model level and then, in a controlled manner, introduced into the production environment. Agile system development encourages Set-Based Design, which entails to test variety of design options. A simulation model that as input receives real time data through dynamic data injection can more accurately establish which design option affords most optimal solution.

With the promises that use of simulation yields with building simulation models on top of applications that are work-in-progress comes a few challenges. First, the *maintainability* of the simulation model due to continuous deployment of new features and the need for more telemetry data. Second, *visualization* or data presentation, i.e., building a dashboard that present the data in an easy fashion and making sense to business and executives, as well as to developers and analyst. These challenges present a great opportunity for profound new research.

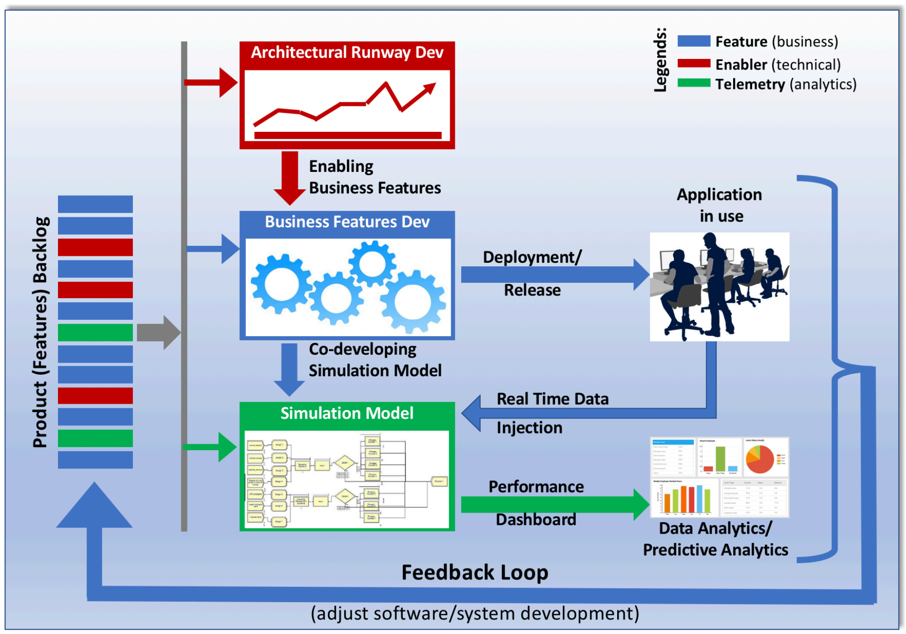

Figure 3: Simulation built on a new software application or system and evolving with the application.

The simulation model maintainability challenge is dominantly a technical challenge, which needs to be addressed through new simulation frameworks and simulation processes. It is similar to the issue of software evolution and keeping documentation updated. Here, one could borrow concepts from *Behavior-Driven Development*, where description of the behavior that a software application has to implement is interwoven into the software itself. That is, a new functionality cannot be developed unless a new behavior is specified and documented, and vice versa, a software development is not complete unless it passes the specified behavior. Such a loop constantly keeps each other, i.e., documentation and the software itself, in correspondence, and, thus, always updated.

The situation is a bit different with addressing visualization. Visualizing the collected data from the application execution in a production environment entails a great deal of data engineering, data science, and data manipulation. Furthermore, this also fosters close collaboration of technical and management people including developers, testers, and business analysts. This collaboration is also known as three amigos, which emphasizes the importance of different perspectives on the system to be built. While nominally they are referred to as three amigos, in fact, more perspectives are needed. One such important perspective is of the user experience or User eXperience (UX) design.

In this regard, simulation elevates to the mainstream of Agile system and software development by assuming the role of collaboration and communication tool as well. In this role, simulation is not merely a technical model or predictive tool, but a vehicle to collaborate, understand, and adjust the system development process.

Finally, there is an input data challenge through dynamic data injection. Data injection challenges are coming from a myriad of factors: the production data is just emerging. In each given moment, the simulation is run on a richer data set than the previous runs. The data injected onto the simulation model from the application is not complete as many of the features are not yet developed. Thus, any optimization of the application will be challenged when new features are implemented in production. Balancing between the production environment injected data, real time data, and mock data needs to be addressed.

6 A DDDAS-BASED ATTACK-RESILIENT FRAMEWORK FOR CRITICAL ENERGY SYSTEM SECURITY (DONG JIN)

In this section, we discuss the emerging role of DDDAS-based methodologies in *cyber resilient* and *secure critical* energy infrastructures and its associated research challenges. The current electric grid is designed to be fault-resilient, i.e., the grid has operating reserves to handle the failure of any single element in the power system, such as a generator or a transmission line. Many power grid applications, however, like state estimation, automatic generation control, and remedial action scheme are not designed to handle failures caused by malicious cyber-attacks. Furthermore, a growing number of new applications are developed and integrated into the power system, which introduces unexpected dependencies in the operation of the various schemes. It is therefore critical to re-examine the existing communication network architecture and application designs with a specific focus on cyber-attack resilience. The challenges arising from the mission-critical and time-critical nature of the critical energy infrastructures are to meet specific requirements, such as continuous availability and functional correctness of the entire system, as well as real-time operations. These challenges call for bold effort.

A DDDAS-based approach can secure the control and communication network of modern power systems. DDDAS is a paradigm that involves dynamically incorporating real-time data into computations in order to steer the measurement and control process of an application system (Darema 2004). The DDDAS-based methodology is used to develop multiple security modules to make the control and communication network of power systems attack-resilient. The high-level framework design is depicted in Figure 4. The correct system behaviors are first modelled (e.g., networks and applications), and then perform verification and testing against desired properties (e.g., security policies and network invariants) with dynamic inputs (e.g., sensor measurement in the device layer, topologies and forwarding tables in the network layer, and control events in the application layer). Violations will indicate vulnerabilities and errors caused by cyber-attacks, misconfigurations or wrong designs. By doing so, the system could be compared before and after a change in configurations to quantify security breaches (e.g., unauthorized access). Additional developments could automate synthesis of the network state and the control logic to enable powerful capabilities, such as immediate response to faults (rather than reaction at human timescales), and self-heal the control and communication networks. The DDDAS framework:

- Contains several dynamic data-driven models that efficiently abstract the network and application behavior (e.g., packet forwarding and power system state). One key feature is that the models are capable of taking dynamic inputs to refine themselves at execution time.

- Includes several algorithms to perform attack-detection and network verification against the desired properties, and then to apply the appropriate mitigation techniques to self-heal the network and restore power system observability and the correct power application state. One key feature is that the algorithms enable dynamic model selection to meet the real-time requirement and certain power system specific constraints and security policies; and
- Enables an efficient feedback control loop to steer the measurement and control process, such as selecting a different set of sensor measurement data for the power system state estimation, or changing the network configuration to mitigate the risks and restore the power system observability.

Research Challenges. The DDDAS-based approach consists of a set of techniques for performing and integrating security and resilience analyses applied at different layers (i.e., data forwarding, network control, and application software) ideally in a real-time and automated fashion. One challenge is to integrate resilience and security evidence gathered in various ways at *different layers* (e.g., formal proofs at one layer, with empirical or simulation-based evaluation at another). Challenges also lie in how to incorporate the *timing uncertainty* into the dynamic data driven models to preserve the consistency as the network state evolve, and how to balance the model complexity (e.g., how many constraints to consider and which security policy and to what details one needs to verify) and the strong real-time requirement of the critical control systems. Additionally, generating solutions to constrained problems is in general more difficult than verifying solutions. It is challenging to *scale solutions* to moderate or large networks on real-world tasks.

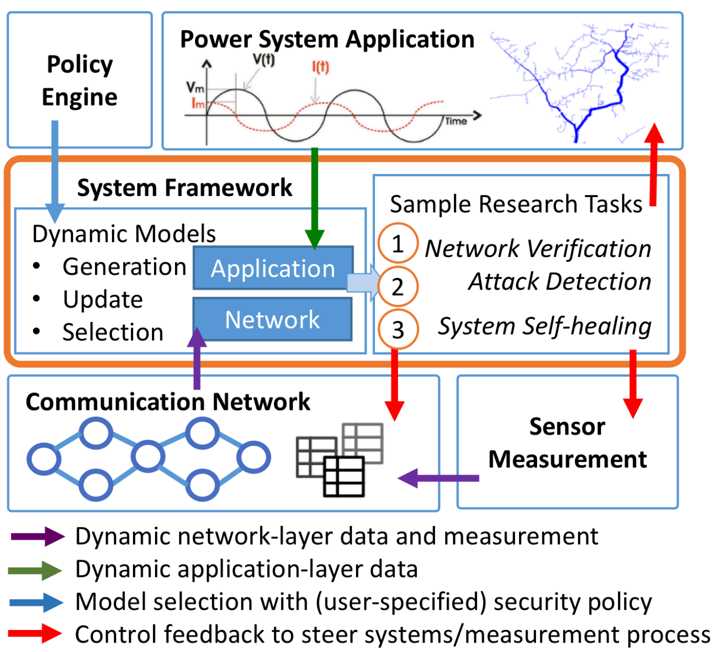

Figure 4: A DDDAS-based Attack-Resilient Framework for Critical Energy System Security.

7 MULTI-SCALE SIMULATION WITH DDDAS FOR BORDER SURVEILLANCE (SEUNGHAN LEE AND YOUNG-JUN SON)

In this section, we discuss how the multi-scale simulation methodologies and physics-based simulation can be applied to handle challenges in border surveillance systems.

7.1 Multi-scale Simulation with DDDAS for Border Surveillance

Border surveillance is a challenging task considering its wide range as well as involvement of various types of sensors. The wide range of surveillance area induces the uncertainty of the information, while various sensor types cause the information heterogeneity problem in information fusion. Wireless sensor network (WSN) technology has been increasingly studied as a solution to these problems. However, the real deployment and operation of WSN in a border area is still highly challenging due to its spatial and temporal scale. To overcome such challenges, multi-scale simulation methodologies have been recently proposed. For example, the Son's group (Lee et al. 2017) has designed and developed an autonomous border surveillance system via unmanned vehicles (UVs) involving two levels of simulation: 1) agent-based simulation (ABS) using Anylogic (top-level) and 2) physics-based simulation (PBS) using Unity3D (low-level). ABS takes charge of generating the initial waypoints of all UVs and monitors the status of the overall surveillance system, while PBS handles the control and communication of UVs in real time. By doing so, the multi-scale modeling approach is enabled, allowing efficient analysis and control for various sensors under the DDDAS philosophy. In the following section, the role of PBS in the border surveillance system is discussed with actual examples of our developed systems.

7.2 The Role of Physics-based Simulation in Border Surveillance

Recently, the research works on PBS have been emerging in different application areas, such as manufacturing, computer science, and surveillance. The necessity of PBS escalates due to the following capabilities: 1) modeling behaviors of mobile sensors, 2) incorporating geographic information system (GIS) information, and 3) enhanced accuracy in human-behavior models. According to a U.S. Customs and Border Protection (CBP) report in 2016, two types of sensors are deployed in the border area: mobile and fixed sensors (U.S. CBP 2016). Mobile sensors include various types of sensors with mobility by being attached to the ground/aerial vehicles. To represent the operational limits of mobile sensors accurately, the PBS can mimic the physics-based phenomena occurring during the operation of mobile sensors. Second, the increased accuracy of GIS information is required in a border surveillance system. There are three types of data in GIS: topography, vegetation, and elevation (Campbell 2011). In our previous surveillance system based on an agent-based simulation tool, the focus was to handle only the elevation data of GIS (Khaleghi et al. 2013). However, by using PBS, the full set of GIS information can be incorporated into the system in real time, which can increase the accuracy of detection and tracking of the target. Third, human behaviors can be modeled in a highly accurate manner. The co-author's group has been working on modeling human behaviors (such as Extended-Belief-Desire-Intention) in different applications, such as transportation, supply chain, and social networks (Lee et al. 2010; Zhao 2008). Since the information of human perception of the environment is an essential input to the model, PBS is essential in providing the input for human behavior models. For example, the field of view (FOV) of an unmanned aerial vehicle (UAV) and line of sights (LOS) of targets are the primary perception results, affecting the tracking and control of targets in the border area. By using PBS with Unity3D, these perception results are well represented enabling DDDAS capabilities in border surveillance. The next section demonstrates the capabilities of PBS with detailed examples.

7.3 Implementation of Surveillance System with PBS via Unity3D

Figure 5 depict the snapshot of the Unity3D (left) of terrain near the US-Mexico border area, and the satellite image taken by Google maps (right). As seen in the figure, all-terrain information can be well incorporated into Unity3D model by maintaining all three types of GIS information. The detailed expression of terrain information enhances the human behavior model in the simulation.

Figure 6 demonstrates the capability of the Unity3D model in generating sensory data (mainly vision data). The first image depicts the actual image captured by a camera attached to a UAV and the second is the simulated image captured by simulated UAV in the simulation, respectively. As both images are at the same altitude (i.e., 7 meters from the ground), the human objects are shown similarly. The third and fourth images demonstrate the images from the camera in a ground vehicle and the simulated camera in the Unity3D model, respectively. These images illustrate and demonstrate the capability and usefulness of PBS as a tool for data collection (e.g. testing algorithms in a border surveillance system) in unreachable areas due to certain environmental constraints.

Figure 5: Terrain in Unity3D Simulation (left) and satellite image in Google map (right).

Figure 6: Images from real camera (first and third) and simulated images (second and forth).

8 CONCLUDING REMARKS

The DDDAS paradigm has had significant impacts in the past decade and can be expected to increase in importance in the years ahead. This paper touches upon a few areas of on-going research, and highlights research challenges and opportunities for DDDAS in the future. A rapidly changing technology landscape coupled with the requirements of new, emerging applications present new challenges and opportunities for future DDDAS deployments. It is apparent that advances in the development of new computational capabilities coupled with new application domains will make DDDAS an increasingly important computational paradigm into the foreseeable future.

The different viewpoints presented in this paper highlight the breadth of research challenges that arise when considering DDDAS. Research challenges span multiple levels of the software stack, including hardware and embedded system design, middleware, analytics, and applications. Common themes include agent simulation, model synchronization, user interaction, and data analytics. Issues span theoretical, algorithmic, and computational aspects. While these authors discuss a few specific application areas, it is clear the DDDAS paradigm spans many different technical domains as highlighted by the work of others in the DDDAS community. Overall, DDDAS supports a broad range of application areas where additional research is needed to fully realize the opportunities afforded by the DDDAS paradigm in modeling, simulation, and run-time execution.

ACKNOWLEDGEMENTS

Research by the authors was supported by the following grants: AFOSR/DDDAS FA9550-17-1-022 (Fujimoto), YIP FA9550-17-1-0240 (Jin), AFOSR/DDDAS FA9550-17-1-0075 (Son).

REFERENCES

Blasch, E. P., E. Bosse, and D. A. Lambert. 2012. *High-Level Information Fusion Management and Systems Design*. Norwood, Massachusetts: Artech House.

Blasch, E., S. Ravela, and A. J. Aved. 2018. *Handbook of Dynamic Data Driven Applications Systems*. Springer.

Biswas, A. and Fujimoto R. M. 2018. “Zero Energy Synchronization of Distributed Simulations”. *Principles of Advanced Discrete Simulation,* 85–96.

Campbell, J. B. and R. H. Wynne. 2011. *Introduction to Remote Sensing*. Guilford Press.

Darema, F. 2004. “Dynamic Data Driven Applications Systems: A New Paradigm for Application Simulations and Measurements”. In *Proceedings of the International Conference on Computational Science*, 662–669. Berlin, Heidelberg: Springer.

DDDAS. 2018. "Dynamic Data Driven Application Systems." Accessed 1st April, 2018. http://1dddas.org/

Dunik, J., O. Straka, M. Simandl, and E. Blasch. 2015. “Random-Point-Based Filters: Analysis and Comparison in Target Tracking”. *IEEE Transactions on Aerospace and Electronic Systems* 51(2):1403–1421.

Feldkamp, N., S. Bergmann, and S. Strassburger. 2017. “Online Analysis of Simulation Data with Stream-based Data Mining”. In *Proceedings of the 2017 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*, May 24th-26th, Singapore, 241–248.

Fujimoto, R. M., C. Bock, W. Chen, E. Page, and J. H. Panchal. 2017. *Research Challenges in Modeling and Simulation for Engineering Complex Systems*. Springer.

HLA Working Group. 2010. *IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA) – Interface Specification*. New York, NY: IEEE.

Imai, S., E. Blasch, A. Galli, F. Lee, and C. A. Varela. 2017. “Airplane Flight Safety Using Error-Tolerant Data Stream Processing”. *IEEE Aerospace and Electronics Systems Magazine*, 32(4):4–17.

Khaleghi, A. M., Xu, D., Wang, Z., Li, M., Lobos, A., Liu, J., and Son, Y. J. 2013. “A DDDAMS-based Planning and Control Framework for Surveillance and Crowd Control via UAVs and UGVs”. *Expert Systems with Applications* 40(18):7168–7183.

Kretz, T., S. Hengst, and V. Roca. 2011. “Calibrating Dynamic Pedestrian Route Choice with an Extended Range Telepresence System”. In *Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops*, Nov 6th-13th, Barcelona, Spain, 166–172.

Lee, S., Y. J. Son, and J. Jin. 2008. “Decision Field Theory Extensions for Behavior Modeling in Dynamic Environment using Bayesian Belief Network”. *Information Sciences* 178(10):2297–2314.

Lee, S., S. Minaeian, Y. Yuan, J. Liu, and Y. J. Son. 2017. “DDDAMS-Based Border Surveillance and Crowd Control via Aerostats, UAVs, and Ground Sensors”. In *Proceedings of 2017 IEEE 2nd International Workshops on Foundations and Applications of Self* Systems (FAS*W)*, September 18th-22nd, Tucson, Arizona, 348–351.

Li, H., K. Sudusinghe, Y. Liu, J. Yoon, M. Van Der Schaar, E. Blasch, and S. S. Bhattacharyya. 2017. “Dynamic, Data-Driven Processing of Multispectral Video Streams”. *IEEE Aerospace and Electronics Systems Magazine* 32(7):50–57.

Leffingwell, D. 2010. *Agile Software Requirements: Lean Requirements Practices for Teams, Programs, and the Enterprise*. 1st ed. Addison-Wesley Professional.

Ma, Y., E. W. M. Lee, and R. K. K. Yuen. 2016. “An Artificial Intelligence-based Approach for Simulating Pedestrian Movement”. *IEEE Transactions on Intelligent Transportation Systems* 17(11):3159–3170.

Mustaffe, N., S. Brailsford, A. Djanatliev, T. Eldabi, M. Kunc, and A. Tolk. 2017. "Purpose and Benefits of Hybrid Simulation: Contributing to the Convergence of its Definition". In *Proceedings of the Winter Simulation Conference*, edited by W. K. V. Chan et al., 1631–1645, Piscataway, New Jersey: IEEE.

Neal, S., R. M. Fujimoto, and M. Hunter. 2016. "Energy Consumption of Data Driven Traffic Simulations". In *Proceedings of the Winter Simulation Conference,* edited by T. Roeder et al., 1119–1130. Piscataway, New Jersey: IEEE.

Neal, S. and R. M. Fujimoto. 2017. "Energy Consumption of HLA Data Distribution Management Approaches". In *Proceedings of the Winter Simulation Conference*, edited by W. K. V. Chan et al., 810–820. Piscataway, New Jersey: IEEE.

Neal, S. and R. M. Fujimoto. 2018 "Power Consumption of Future Event List Implementations in Discrete Event Simulations". In *Proceedings of Annual Simulation Symposium*, April 15th-18th, Baltimore, Maryland, 9.

Nelson, B. L. 2017. "WSC 2067: Where Are the Chances?". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, et al., 1, Piscataway, New Jersey: IEEE.

Picciano, B. 2014. *Why Big Data Is The New Natural Resource*. Forbes BrandVoice, accessed Jun 30th, 2014.

Taylor, S. J. E., A. Khan, K. L. Morse, A. Tolk, L. Yilmza, J. Zander, and P. J. Mosterman. 2015. "Grand Challenges for Modeling and Simulation: Simulation Everywhere – from Cyberinfrastructure to Clouds to Citizens". *Simulation* 97(7):648–665.

Snidaro, L., J. Garcia Herrero, J. Llinas, and E. Blasch. 2016. *Context-Enhanced Information Fusion*: *Boosting Real-World Performance with Domain Knowledge*, Springer.

Tolk, A. 2015. "The Next Generation of Modeling and Simulation: Integrating Big Data and Deep Learning". In *Proceedings of the Conference on Summer Computer Simulation*, July 26th-29th, Illinois, Chicago, 1–8.

U.S. Customs and Border Protection. 2016. CBP Annual Report. https://www.cbp.gov/sites/default/files/assets/documents/2017-Mar/FY-2016-CBP-PAR-508C.pdf, accessed 20.05.2018.

Viswananthan, V., M. Lees, and P. M. A. Slot. 2017. "The Influence of Memory on Indoor Environment Exploration: A Numerical Study". *Behavior Research Methods* 48(2):621–639.

Wang, M., and X. Hu. 2015. "Data Assimilation in Agent Based Simulation of Smart Environment Using Particle Filters". *Simulation Modelling Practice and Theory* 56(80):36–54.

Ward, J. A., A. J. Evans, and N. S. Malleson. 2016. "Dynamic Calibration of Agent-based Models Using Data Assimilation". *Royal Society Open Science* 3(4):150703.

Zhao, X. and Y. J. Son. 2008. "Extended BDI Framework for Modelling Human Decision-Making in Complex Automated Manufacturing Systems". *International Journal of Modelling and Simulation* 28(3):347–356.

Zheng, Y., E. Blasch, and Z. Liu. 2018. *Multispectral Image Fusion and Colorization*, SPIE.

Zhong, J., W. Cai, M. H. Lees, L. Luo, and N. Hu. 2017. "Automatic Model Construction for the Behaviour of Human Crowds". *Applied Soft Computing* 56(7):368–378.

AUTHOR BIOGRAPHIES

RICHARD FUJIMOTO is a Regents' Professor in the School of Computational Science and Engineering at the Georgia Institute of Technology. He received a Ph.D. in Computer Science & Electrical Engineering from the University of California-Berkeley in 1983. He has been an active researcher in the parallel and distributed simulation field since 1985, and has authored or co-authored 3 books and over 250 technical papers on this subject including several award winning publications. He led the definition of the time management services for the High Level Architecture for modeling and

simulation (IEEE standard 1516). He received the ACM Distinguished Contributions in Modeling and Simulation Award in 2013. His e-mail address is fujimoto@cc.gatech.edu.

JOSEPH BARJIS is Director of Agile Transformation at PricewaterhouseCoopers and Adjunct Professor at San Jose State University. Dr. Barjis research works are published in 100+ peer refereed outlets (over 25 journal articles and editorials; 17 book chapters; 18 edited proceedings books, 10 Special Issues Editor; 60+ papers in peer refereed conference; published in high impact journals), including a number of Best Paper Awards. He is co-founder of the emerging discipline of Enterprise Engineering and founder of an international workshop in the field, now on its 14th annual edition. Dr. Barjis has 25+ years of technology and business acumen in consulting, R&D, training, and teaching. Thought leader and expert in Agile, Systems Engineering, Enterprise Engineering, Organization Design, Predictive Analytics, Software Development, and Information Systems Design. He has been internationally invited for talks, scientific seminars, or lectures. He worked from top research universities to top consulting companies. His email address JBarjis@gmail.com.

ERIK BLASCH is a program officer at the Air Force of Scientifc Research. He received his BS from MIT, multiple MS degrees, and PhD from Wright State University. He has compiled 750+ papers, 21 patents, and 25 tutoirals. He is a member of ISIF, AIAA Associate Fellow, SPIE Fellow, and IEEE Fellow. His email address is erik.blasch.1@us.af.mil.

WENTONG CAI is a Professor in the School of Computer Engineering at Nanyang Technological University, Singapore. He received his Ph.D. in Computer Science from University of Exeter (UK) in 1991. His expertise is mainly in the areas of Modeling and Simulation and Parallel and Distributed Computing. He has published extensively in these areas and has received a number of best paper awards at the international conferences for his research in distributed simulation. He is an associate editor of the *ACM Transactions on Modeling and Computer Simulation* (TOMACS) and an editor of the *Future Generation Computer Systems* (FGCS). His email address is aswtcai@ntu.edu.sg.

DONG JIN is an Assistant Professor in the Computer Science Department at the Illinois Institute of Technology. He obtained his Ph.D. in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign. His research interests include simulation modeling and analysis, trustworthy cyber-physical critical infrastructures, software-defined networking, and cyber-security. He is a recipient of the Air Force Office of Scientific Research (AFOSR) Young Investigator Program (YIP) award. His email address is dong.jin@iit.edu.

SEUNGHAN LEE is a Ph.D. candidate in Systems and Industrial Engineering Department at University of Arizona. His research interests are modeling and analysis of a hierarchical decision-making process in social networks based on simulation and optimization. Major application areas of his research are 1) social sensing and delivery for disaster management and 2) dynamic-data-driven adaptive multi-scale simulation (DDDAMS) for border surveillance. He has received the ISERC Conference Best Paper Awards (2016 and 2018). He can be reached at shlee17@email.arizona.edu.

YOUNG-JUN SON is a Professor and Head of Systems and Industrial Engineering at the University of Arizona. His research focuses on the coordination of a multi-scale, networked-federated simulation and decision model needed for design and control in manufacturing enterprise, renewable energy network, homeland security, agricultural supply network, and social network. He is an IISE Fellow and has received several research awards such as the SME 2004 Outstanding Young Manufacturing Engineer Award, the IISE 2005 Out- standing Young Industrial Engineer Award, the ISERC Conference Best Paper Awards (2005, 2008, 2009, 2016, 2018), and Best Paper of the Year Award in 2007 from IJIE. He can be reached at son@sie.arizona.edu.