Request for Information (RFI): SLAM

System-Level Application Modeling (SLAM)

IARPA-RFI-19-08

The Intelligence Advanced Research Projects Activity (IARPA) is seeking information on research efforts in the area of future modeling and simulation (hereafter “ModSim”) research for large-scale computational and data-analytic applications. This request for information (RFI) is issued solely for information-gathering and planning purposes; this RFI does not constitute a formal solicitation for proposals. The following sections of this announcement contain details about the scope of technical efforts of interest, along with instructions for submitting responses.

Background & Scope

Modeling and simulation (ModSim) is essential to the co-design of new system architectures (hardware and software) and applications. It provides a cost-effective way to explore new designs, to examine the impact of excursions from a baseline design and to optimize the hardware and software before a final build.

The ModSim challenges posed by escalating system and application complexity are many and include: application execution, high-performance data movement, data management, computation scheduling, and representation of system characteristics. An application could involve a mix of streaming data and file system-based bulk data. This additional challenges of heterogeneous data sources make modeling the execution of an application an even more important, but complicated effort.

As the complexity of high-performance computing systems grows in response to demand for greater capacity and capabilities, the challenges of modeling the execution of an application on these systems are rapidly escalating. Current and future system architectures many-core designs, deeper memory hierarchies with non-volatile technologies and burst buffers, novel data storage components, processing near memory, increased on-node heterogeneity, novel networking approaches, and extreme system-level parallelism. A system could be composed of a variety of disparate computers, storage resources, networks, and sources of large-scale data. All of these factors are driving the demand for balanced performance in future applications executing not only on increasingly complex, heterogeneous systems, but also under increasingly severe system and facility power constraints while simultaneously providing high levels of security and resiliency. The complexity of hardware-software interactions is further exacerbated by the need to satisfy the convergence of high-performance computing and data-analytics applications and the associated challenges of dealing with irregular data structures and dynamically-changing environments.

Providing insight and understanding into how to use these systems effectively and how to design and optimize future systems is vital for the computing ecosystem, from edge computing to extreme-scale
heterogeneous systems. What is required are unified models that can act on dynamic information about the hardware system and running applications; predict power, performance, and resiliency; and formalize the effects of trade-offs as changes occur. The models that capture these interactions guide hardware design, system architecture, algorithm design, and application and data movement optimization. Moreover, the models are critical to future system success. However, progress in ModSim depends critically on overcoming the vast abstraction gaps among hardware, runtime, and application domains. Future ModSim methods must capture all these areas into integrated models.

This RFI seeks to identify advances in the following research disciplines:

- **High-Fidelity ModSim Techniques.** New high-fidelity models of complex and heterogeneous architectures designed to execute large scale data-centric applications, under dynamic conditions of system performance and optimization require new modeling methods and tools. New methodologies must be flexible and low in computational demands. Simulation fidelity and accuracy must be high and provable or, at a minimum, consistent over a diversity of architectures and application workloads. Of particular interest is near real-time high fidelity simulation.

- **ModSim of Artificial Intelligence Systems and Applications and Machine Learning as a Method of ModSim.** Artificial Intelligence (AI), in general, and Machine Learning (ML), in particular, have emerged as essential application drivers in all forms of computing, including large-scale data analytics and numerically-intensive computations. The trend’s impact extends beyond the nature of architectures optimized for executing an ML workload. It also points toward applying AI/ML techniques as ModSim methodologies to support a range of systems (including but not limited to AI-centric systems).

- **Unified Modeling of Performance, Power, and Resilience.** As systems scale in size and technology moves toward highly-resilient and energy-efficient system designs, it is critically important to develop new integrated methods that capture performance, power/energy consumption, and resilience while considering the impacts of thermal effects and power constraints. For example, software-level resilience techniques, including algorithm-based fault tolerance, checkpointing, and architecture-level optimizations may be required to operate at low supply voltages. Such models would enable quantification of the trade-offs between performance, power/energy consumption, and reliability.

- **System-Level Modeling and Simulation.** As defined here, a “system-level model” would involve integrative ModSim technologies to predict performance and energy consumption of the whole system while the system is executing an application of interest. It is highly desirable to be able to design ahead and optimize before and after implementing the whole system using ModSim. The integration of ModSim methods must be over the subsystems and their interfaces and incorporate the ability to “dial-in” resolution and fidelity, depending on the simulation’s intended use. System-level modeling should cover diverse applications of interest with different complexity and workload characteristics. Such an integrated simulation could require mixing and matching methods based on different modeling approaches (e.g., analytical and statistical).
• **Modeling Irregularity in Applications.** Dynamic behavior stemming from both applications and system adaptivity will soon become an essential aspect in normal system operation. As execution models move toward more dynamic task-oriented models, predicting application performance in advance becomes even more challenging due to migratable work units and automatic load balancing mechanisms. Moreover, capabilities such as dynamic power steering, thermal throttling, and process variation result in performance heterogeneity across the system, rendering even carefully initially balanced workloads suboptimal. Capturing this variability is fundamental to advancing state of the art in ModSim techniques.

• **Dynamic Modeling.** ModSim methods that can quantitatively and accurately capture dynamic and adaptive applications and systems behavior are needed. Current leading-edge modeling techniques produce static models that target traditional scientific workloads and are therefore inadequate for addressing dynamic behavior. Existing proven methods must be extended to predict performance and energy consumption of applications whose behavior is input-dependent and dynamic throughout execution. Establishing a dynamic, adaptive modeling methodology would be a crucial enabling technology for efficient and productive future large-scale systems utilization.

As IARPA aims to chart the future of impactful ModSim environments, it is imperative that submitted responses cover a spectrum of ModSim research disciplines. Responses to this RFI should answer the following questions:

1. Identify candidate research efforts for future ModSim environments.
2. Describe a technical approach for modeling the execution of data centric application on large scale computer systems, including the limitations of the methodology.
3. Explain how the implementation of the proposed capability is achievable at the current stage of technology development or what future developments will be needed to achieve implementation.

**Preparation Instructions to Respondents**

IARPA requests that respondents submit ideas related to this topic for use by the Government for information gathering and planning purposes. IARPA requests that submittals briefly and clearly describe the potential approach or concept, outline critical technical issues/obstacles, describe how the approach may address those issues/obstacles, and comment on the expected performance and robustness of the proposed approach. If appropriate, respondents also may choose to provide a non-proprietary rough order of magnitude (ROM) estimate regarding what their approaches might require in terms of funding and other resources to achieve a defined milestone in model fidelity in one or more years. This Request for Information contains all of the information required to submit a response. No additional forms, kits, or other materials are needed.
IARPA appreciates responses from all capable and qualified sources from within and outside of the United States. Because IARPA is interested in an integrated approach, responses from teams with complementary areas of expertise are encouraged.

Responses should adhere to the following formatting requirements:

1. A one-page cover sheet that identifies the title, organization(s), and respondent’s technical and administrative points of contact—including names, addresses, phone and fax numbers, and email addresses of all co-authors—clearly indicating its association with RFI-19-08.

2. A substantive, focused, half-page (0.5) executive summary

3. A description of the technical challenges and suggested approach(es) (limited to eight pages in minimum 12-point Times New Roman font, appropriate for single-sided, single-spaced 8.5 by 11 inch paper, with 1-inch margins)

4. Optionally, a single overview briefing chart graphically depicting the key ideas.

Submission Instructions to Respondents

Responses to this RFI are due no later than 4:00 p.m., Eastern Time, on July 29. All submissions must be electronically submitted to dni-iarpa-rfi-19-08@iarpa.gov as a PDF document. Inquiries to this RFI must be submitted to dni-iarpa-rfi-19-08@iarpa.gov. Do not send questions with proprietary content. No telephone inquiries will be accepted.

Disclaimers and Important Notes

This is an RFI issued solely for information and planning purposes and does not constitute a solicitation. Respondents are advised that IARPA is under no obligation to acknowledge receipt of the information received or provide feedback to respondents with respect to any information submitted under this RFI.

Responses to this notice are not offers and cannot be accepted by the Government to form a binding contract. Respondents are solely responsible for all expenses associated with responding to this RFI. IARPA will not provide reimbursement for costs incurred in responding to this RFI. It is the respondent's responsibility to ensure that the submitted material has been approved for public release by the information owner.

The Government does not intend to award a contract on the basis of this RFI or to otherwise pay for the information solicited, nor is the Government obligated to issue a solicitation based on responses received. Neither proprietary nor classified concepts nor information should be included in the submittal. Input on technical aspects of the responses may be solicited by IARPA from non-Government consultants/experts who are bound by appropriate non-disclosure requirements.

Contracting Office Address:
Office of the Director of National Intelligence

Intelligence Advanced Research Projects Activity

Washington, District of Columbia 20511

United States

**Primary Point of Contact:**

William J. Harrod

Intelligence Advanced Research Projects Activity

dni iarpa rfi 19 08 iarpa gov