LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-664518

Situational Awareness - Cross-cutting Capability for LLNL Global Security Programs

J. M. Brase, E. M. McKinzie

November 19, 2014

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Situational Awareness – Cross-cutting Capability for LLNL Global Security Programs

Today's national security priorities require understanding of the activities and behaviors of very complex, interconnected systems – the national power grid, large-scale information networks, financial, and organizational networks. Decision-makers need predictive analytics that can provide timely answers to questions like

- Is the activity we're seeing normal?
- Do we see indications of threat activities?
- What will happen if we change this system element?
- How certain is this conclusion?

Rapidly expanding access to data on the structure and activity of these systems, combined with continually growing computing power, is enabling a new set of tools to address these questions in a wide variety of application areas from business to national security. LLNL combines leading edge R&D in machine learning and graph analytics with the country's leading supercomputing skills and tools to create unique situational awareness analytics.

In national security applications such as WMD proliferation and terrorism, deep subject-matter expertise in the related science disciplines plays a critical role in the success of the analysis. Integration of this deep domain knowledge into automated computing tools forms the basis for LLNL's approach to situational awareness.

Laboratory capabilities and resources

LLNL provides longstanding leadership in national security science and technology areas. Programs in nuclear security and nonproliferation, counter-terrorism, and homeland security have developed deep expertise in areas including nuclear detection and analysis, high explosives, chemical and biological forensics, and cyber and information operations. Expertise from all these areas is integrated into LLNL situational awareness and analytics systems. Ultimately this expertise resides in LLNL's staff of 2000 scientists and engineers, half at the PhD level, that are collectively responsible for awards ranging from three Nobel Prizes to 148 R&D 100 technology awards over the past twenty years.

Large-scale computing and data analytics is central to essentially all LLNL programs and capabilities. LLNL partners with the world's leading computing manufacturers to develop and operate some of the world's top-ranked high-performance computers. The 2013 Gordon Bell Prize for the year's leading high performance computing achievement was awarded to a collaboration based on LLNL's 20

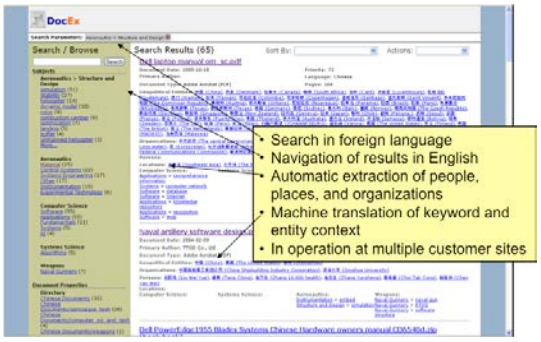
petaflop Sequoia system. We continue to invest in leading-edge technology systems such as Catalyst – a new supercomputing cluster focused on large-scale data analytics and developed jointly by LLNL, Intel, and Cray.

LLNL Situational Awareness tools and programs

LLNL has developed a variety tools and systems to enhance situational awareness focusing on WMD and cyber application areas. Using these tools our Global Security programs answer questions and provide analysis for national programs in the IC, the DOD, and DHS. This direct connection between technology developers and operational national security analysts has greatly enhanced the usability and impact of the tools. Specific LLNL tools and capabilities include:

- The Counter-proliferation Analysis and Planning Systems (CAPS) uses its Trinidad document processing tools to build situational awareness data sets that support WMD detection and response planning. Hadoop-based natural language tools, coupled with SME-developed WMD process models, search and analyze up a billion on-line documents at all levels of classification.

- The Biodefense Knowledge Center (BKC) uses Web search and content extraction from technical document streams to support biosecurity threat analysis for DHS customers.

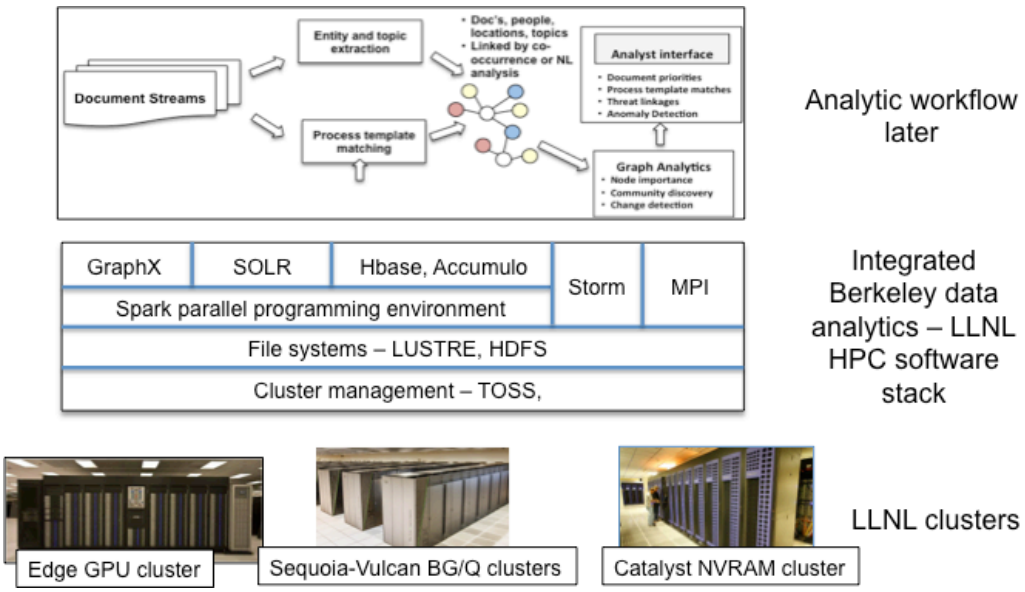


- The Information Operations and Analysis (IOA) Program develops and fields advanced tools and methods for cybersecurity. Large-scale network mapping, machine learning, and graph analytics are used to understand activities in computer networks. Faceted information navigation interfaces provide tools for analysts to organize and prioritize large document data sets for cyber applications.

A computing framework for high–performance analytics

Bringing state-of-the-art high-performance computing capability to data analytics will enable new approaches to integrating science-based models and simulations and allow analytics to scale to increasing data volumes and model complexities. LLNL is working with UC Berkeley's AMPLAB, Stanford University, and Auburn University to integrate leading industry approaches to data analytics software with

HPC computing systems and software environments. Ultimately this integration will allow analysis workflows developed in a commodity cluster environment using the best new toolsets to port directly into an HPC environment. It also enables new approaches to integrating high-performance simulation models into the analytics workflow – an emerging need in analyzing complex systems.



LLNL's roadmap for situational awareness R&D

Today, LLNL situational awareness capabilities are centered on scalable natural language technologies feeding both unstructured text and structured entity/transaction data to large-scale relational databases supporting analyst search functions. We are continuing to invest LDRD and other funding to develop new approaches, methods, and tools in several directions:

- New approaches to fully incorporating SME-based science models into analytic tools. New natural language approaches coupled with new data-driven simulation capabilities will lead to unique science-based data analytics.
- Expanded knowledge representations, particularly focused on semantic graphs, will enable improved performance with very low signal-to-noise data as well as new analytic approaches.
- Extended data modalities will allow programs to integrate image, video, and audio data into systems that currently handle only document and structured data.

- New analytics will extend beyond our current search-centered capability to tools that produce causal mechanistic models for activities, detect anomalous changes in time, and that automatically summarize reporting for analysts.
- Finally, new hybrid computer architectures that bring together commodity data-intensive approaches like Hadoop and Spark with high-performance memory and communication architectures (e.g. the Catalyst architecture discussed above) will extend usable situational awareness capability to truly global scales.

Nearly all our situational awareness and analytics R&D is done in collaborations with leading academic and industry partners. We work with universities ranging from Stanford and UC Berkeley to Georgetown, Carnegie-Mellon, and Auburn. We work with the Center for Urban Science and Progress at New York University to develop analytics and simulations to be applied to newly emerging data on behaviors and activities in cities. In partnership with IBM, LLNL's Sequoia computer system has been ranked number one in graph analysis performance for three consecutive years. These partnerships play an important role in keeping LLNL at the forefront of situational awareness R&D and in sustaining a vital and growing workforce.

LLNL Point of Contact

Eric McKinzie, Global Security Directorate

mckinzie1@llnl.gov

925-422-9244