OAK RIDGE NATIONAL LABORATORY FACT SHEET
MANAGED BY UT-BATTELLE FOR THE DEPARTMENT OF ENERGY

Forward Analysis and Alerting

Analyzing Data Where It Lives

The Problem

For intelligence analysts to prevent or respond to threats, they must have a comprehensive ability to track threats and identify and exploit key relationships and alliances. Unlike law enforcers, who routinely gather evidence to support an investigation after a crime has been committed, analysts gather information to try to stop the crime before it *might* happen.

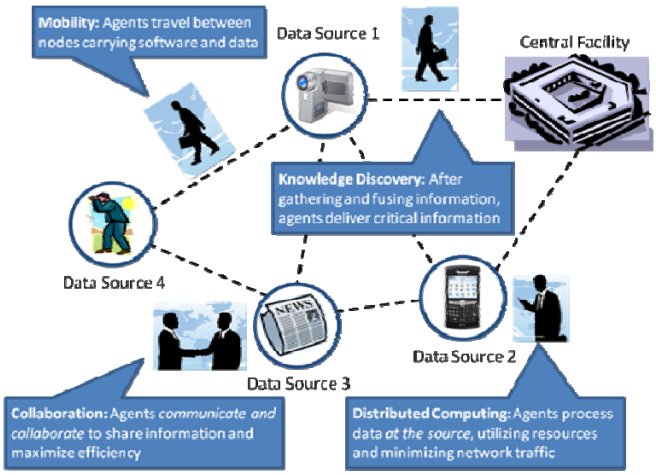
The truly useful tidbits of information that could help analysts are surrounded by a sea of diverse and apparently conflicting information. Often these tidbits are oblique and represent a highly fragmented, partial view of the truth. The process of intelligence analysis is akin to taking random snapshots of a panoramic ocean view through a soda straw. The result is many small fragments of the seascape that could support a broad range of conclusions about the overall picture. Separating the useful clues from the surrounding sea and creating a whole picture is not trivial.

This is often exacerbated by the data owners' reluctance to share information, in most cases for very important reasons. So the question is how you can do this type of analysis when you are only allowed to access the information but not load the information into your system. We propose a grand challenge to address these issues, *Forward Analysis and Alerting*. This is a distributed data mining capability utilizing software agents that are aware, i.e., they understand what kind of data they need to complete their mission, where they need to go to find that data, how to do the appropriate analytics, and who should be alerted as to the results of the analytics.

The Concept

We address two interdependent challenges in enabling access to information when such information is dynamic (i.e., streaming and real-time), distributed, and voluminous. First, we pose and will address the Wide-Area Dynamic-Data Instantiation challenge: given a search query specification (i.e., an event or data of interest) and a set of data-descriptors, how does a system automatically instantiate a real-time data overlay network to collect, organize, semantically-tag, retrieve the data, and deliver it to the appropriate person(s)? And, second, how does a distributed processing capability based on a software agent paradigm evaluate (or "forward analyze") instantiated streaming data flows in a heterogeneous, distributed, low-bandwidth environment to support the created real-time overlays in a secure environment?

Quite simply, we propose to send software (and analysis commands) to the information, and return just the information that is needed. Going past the traditional approaches that are based on centralizing data-access, improving the search technology, or at most, retrieving the data from distributed repositories, we propose a new method for meeting user needs in such Exabyte environments. Distributed data mining algorithms can provide answers faster by taking advantage of the local processing capability of the data-generating devices. Thus, *forward analysis*, or pushing computation to data, would enable real-time analysis and alerting.



The Technical Solution

We build upon state-of-the-art research and development in event processing and develop a novel capability that integrates event-processing with distributed processing using software agents for autonomous operations in heterogeneous environments. Agent-based software enhances distributed data mining and forward analysis by leveraging the autonomous nature of software agents. Mobile software agents hop across systems to carry out non-predefined tasks. The software agent-based distributed analysis has the capability to analyze data as well as to collaborate with other software agents toward a goal-driven solution, i.e., the software agent is aware. It understands what kind of data it needs, where it needs to go to find that data, how to do the appropriate analytics, and who should be alerted to the results of the analytics.

Point of Contact:

Rick Lusk
Computation Sciences & Engineering Division
Oak Ridge National Laboratory
P.O. Box 2008, MS 6086
Oak Ridge, TN 37831-6086
Phone: 865-574-8864
E-mail: luskrm@ornl.gov