Semantic Information Channelling

Tatiana Kichkaylo, Michael D. Orosz, Robert Neches (presenter)

Presented at the Undersea Distributed Networked Systems Conference, Newport, Rhode Island, February 13-15, 2007, session on Information Space Management

A key challenge to the Undersea DNS is managing the collection, processing, presentation, and archiving of information required for continuous 24/7 operations.

Modern sensor and reconnaissance technologies produce huge amounts of data. Communication and storage systems can deliver this data to any point in the network and store it for historical analysis. However, the very amount of the available information decreases its usability. To be useful, relevant and only relevant information needs to be delivered to the right people at the right time, and potentially relevant information (including archived data) should be easily searchable. On the other hand, excessively proactive pruning of data may prevent crucial information from reaching the right people.

Relevance of a particular data item to a situation and person depends on multiple factors, such as the role of the individual in the organization, accuracy and freshness of the data, pace of the events, and potential implications of the data. Given the variety of information sources and information uses, it is unlikely that a rigid information distribution schema will work well in all cases. Instead, a dynamically configurable information distribution framework is needed. This framework should support both push (for delivering fresh data with significant implications to the user) and pull (for evaluation of alternatives and historical analysis) modes.

At the University of Southern California's Information Sciences Institute (ISI), we are developing technologies that address many of the requirements listed above. The Semantic Network infrastructure we are developing as part of the Risk Analysis Workbench (RAW) project provides a framework for dynamically linking information. In this infrastructure, every piece of data and information source represent nodes of the semantic network. Nodes are tagged with a set of semantic properties. Semantic links (i.e. links with "meaning" attached to them) connect the nodes. For example, a UUV and an enemy submarine are nodes, and so is a record of the UUV detecting the submarine. The detection record is connected by "source" and "object" links to the UUV and the submarine respectively. Further, the UUV might have "location" and "reliability" properties associated with it.

Because semantic tags and links are structured data, rather than free text, their use provides two key advantages over rigid conventional publish-and-subscribe approaches. First, they enforce a degree of organization which still permits users to define new topics

but without the clutter of mismatches which arises when plain text is used unsystematically. Second, the structure enables use of inference mechanisms, allowing users to find information items that are logically implied even if not tagged with the exact terms they used to describe their interest.

The information space in RAW can consist of multiple servers and information sources, which may be distributed both geographically and organizationally. The Semantic Network can be traversed in various ways, allowing different users to have different views of the same information space depending on their organizational role and the task being performed. For example, a Situation Assessment Officer may start with the detection record, and traverse the network to obtain the data on the quality of the record (based on the reliability of the UUV sensors) and on the submarine (based on detection records from other sensor devices located in the same area as the UUV).

Semantic queries can be submitted to the RAW network to obtain (pull) information relevant for the task, such as information about available resources capable of obtaining additional information on the detected submarine or a set of previously recorded cases similar to the situation at hand. Queries can also be posted as subscriptions, providing basis for the push path of information distribution. Subscriptions can be dynamically generated based on the task or the role of an individual in the organization to ensure that the relevant information is delivered to the right people. For example, a Situation Assessment Officer on duty at a future submarine attack center can be automatically subscribed to enemy vessel detection records from all sensors located in a particular part of the ocean.

The RAW infrastructure is open, allowing both information sources and information sinks to be dynamically added to and removed from the system. In particular, external automated analytical systems can be linked to RAW to augment the information space with information such as possible implications of a data item or reliability of a particular data source (e.g. sensor) based on historical data. For example, if the number of false detections is high, an analytical system may be inserted into the notification path to let through only detection records with high reliability or those corroborated by several sensors.

Bios

Dr. Robert Neches heads the Distributed Scalable Systems Division at the University of Southern California Information Sciences Institute (USC ISI). His interests span distributed systems engineering, information management, intelligent human-computer interaction, computer-supported cooperative work, resource management, and decision support.

Dr. Tatiana Kichkaylo is a Computer Scientist in the Distributed Scalable Systems Division at USC ISI. Her interests span planning, scheduling, search techniques for large-scale problems, autonomous systems, intelligent environments, flexibility-performance tradeoffs in real-world systems, knowledge modeling, decision support, web/grid services, and the semantic web.

Dr. Michael Orosz is a Project Leader in the Distributed Scalable Systems Division at USC ISI, responsible for a wide range of large-scale systems. His interests span human-computer interfaces in mission critical environments, expert systems, and real-time data acquisition systems.