New system will detect insider threats from massive data sets

November 10 2011

To develop new approaches for identifying “insider threats” before an incident occurs, Georgia Tech researchers will have access to massive data sets collected from operational environments where individuals have explicitly agreed to be monitored. The information will include electronically recorded activities, such as computer logins, emails, instant messages and file transfers. Credit: Georgia Tech/Rick Robinson

When a soldier in good mental health becomes homicidal or a government employee abuses access privileges to share classified information, we often wonder why no one saw it coming. When looking through the evidence after the fact, a trail often exists that, had it been noticed, could have possibly provided enough time to intervene and prevent an incident.

With support from the Defense Advanced Research Projects Agency

(DARPA) and the Army Research Office, researchers at the Georgia Institute of Technology are collaborating with scientists from four other organizations to develop new approaches for identifying these "insider threats" before an incident occurs. The two-year, $9 million project will create a suite of algorithms that can detect multiple types of insider threats by analyzing massive amounts of data -- including email, text messages and file transfers -- for unusual activity.

The project is being led by Science Applications International Corporation (SAIC) and also includes researchers from Oregon State University, the University of Massachusetts and Carnegie Mellon University.

"Analysts looking at the electronically recorded activities of employees within government or defense contracting organizations for anomalous behaviors may now have the bandwidth to investigate five anomalies per day out of thousands of possibilities. Our goal is to develop a system that will provide analysts for the first time a very short, ranked list of unexplained events that should be further investigated," said project co-principal investigator David A. Bader, a professor with a joint appointment in the Georgia Tech School of Computational Science and Engineering and the Georgia Tech Research Institute (GTRI).

Under the contract, the researchers will leverage a combination of massively scalable graph-processing algorithms, advanced statistical anomaly detection methods and knowledge-based relational machine learning algorithms to create a prototype Anomaly Detection at Multiple Scales (ADAMS) system. The system could revolutionize the capabilities of counter-intelligence community operators to identify and prioritize potential malicious insider threats against a background of everyday cyber network activity.

The research team will have access to massive data sets collected from

operational environments where individuals have explicitly agreed to be monitored. The information will include electronically recorded activities, such as computer logins, emails, instant messages and file transfers. The ADAMS system will be capable of pulling these terabytes of data together and using novel algorithms to quickly analyze the information to discover anomalies.

"We need to bring together high-performance computing, algorithms and systems on an unprecedented scale because we're collecting a massive amount of information in real time for a long period of time," explained Bader. "We are further challenged because we are capturing the information at different rates -- keystroke information is collected at very rapid rates and other information, such as file transfers, is collected at slower rates."

In addition to Bader, other Georgia Tech researchers supporting key components of this program include School of Interactive Computing professor Irfan Essa, School of Computational Science and Engineering associate professor Edmond Chow, GTRI principal research engineers Lora Weiss and Fred Wright, GTRI senior research scientist Richard Boyd, and GTRI research scientists Joshua L. Davis and Erica Briscoe.

"We look forward to working with DARPA and our academic partners to develop a prototype ADAMS system that can detect anomalies in massive data sets that can translate to significant, often critical, actionable insider threat information across a wide variety of application domains," said John Fratamico, SAIC senior vice president and business unit general manager.

Provided by Georgia Institute of Technology

Citation: New system will detect insider threats from massive data sets (2011, November 10)