GSN: Government Security News

How intelligence analysts ‘connect the dots’ to thwart terrorist plots

Fri, 2011-09-09 12:02 PM
By: Jacob Goodwin

U.S. soldiers seize documents in Iraq

The recent stirrings within the intelligence community that have led to fears of another possible terrorist attack in New York City or Washington, DC, on September 11th supposedly originated in Pakistan, perhaps among the computer files and papers gathered up in Osama bin Laden’s secret compound in Abbottabad the night he was killed last May.

Inevitably, such talk leads to the familiar question, “Will the U.S. intelligence community be able to *‘connect the dots’* that might lead to the perpetrators of such a potential terrorist plot?”

Just what does it mean to “connect the dots”? How many dots are we talking about? And what technological advances have been made in recent years that have empowered the CIA -- and other intelligence agencies that fall under the purview of the Director of National Intelligence (DNI) -- to connect those dots in a faster and more powerful manner?

The answers to those questions might not provide much grist for a Hollywood thriller, but they could make for a compelling PhD. thesis in mathematics, link-analysis, logic and decision-making theory.

It all comes down to the capabilities of the *databases* that are used to discover relationships between millions of different people around the world -- and the *hundreds* of terrorists lurking secretly among those millions of people.

Here’s a hypothetical example that might make this clear. Suppose we take Osama bin Laden as the starting point for one chain of terrorist relationships that have spread over the years from his Al Qaeda training camps and personal residences in Afghanistan and Pakistan, to allies, friends, employees and bona fide terrorists operating throughout the world.

The intelligence challenge is to identify *everybody* that bin Laden has ever communicated with, or dealt with in any way; and then to identify everybody that each of those bin Laden contacts (at “one degree of separation”) ever dealt with in any way, and then everybody at two degrees of separation that they, in turn, ever dealt with in any way, and then at three degrees of separation, etc., etc., etc. To identify all of these relationships, intelligence analysts might examine every telephone number ever dialed by bin Laden on a captured cell phone, or every bank account he ever used (to see whether anyone else was ever listed as an authorized user of the same account), or every person he ever employed, or trained or met with.

Each of those contacts might represent an intelligence “dot.” Remember: we’re trying to “connect the dots.” Suppose, through years of conscientious fact-gathering, U.S. and allied intelligence agencies had identified *one thousand* people that had a direct relationship of some sort with bin Laden himself.

Now, the same fact-gathering process -- checking telephone records, airline tickets, employment histories, home addresses, intercepted conversations, email messages, etc., etc. – might be applied to each of those one thousand bin Laden contacts to see who *they* know. Needless to say, this will expand the entire pool of contacts to hundreds of thousands -- or millions -- of individuals.

By the time the pool has been widened to include the “fourth degree of separation” or the fifth or the sixth, the total population being examined could, conceivably, number in the tens of millions of individuals. The magnitude of the task of crunching all of this data has become one of the major obstacles to successfully “connecting the dots.”

Let’s push my hypothetical example a little further. Suppose a tip is picked up in the mountainous, lawless region of Pakistan suggesting that an unknown Pakistani-American bank teller living and working in Lincoln, Nebraska -- we’ll call him “Ali” – might be involved in a murderous terrorist plot that is targeting Washington, DC, in the next few weeks. A U.S. intelligence analyst assigned to run this lead to the ground might want to gather as much personal data about “Ali” as possible, and then look for “links” that would suggest a relationship with one or more of the tens of millions of individuals who have already been linked, in one way or another, to Osama bin Laden. That task requires an enormous amount of data-processing power, which until recently has simply not been available.

To better understand the challenges of *adding* millions of pieces of information to ever-expanding intelligence databases and then *crunching* that mountain of data to yield meaningful results, *Government Security News* recently spoke with John “Jay” Jarrell, president and CEO of Objectivity, Inc., of Sunnyvale, CA, which has developed the “enabling technology,” which it calls *InfiniteGraph*, for a database truly on steroids.

In the “old days” (meaning a few years ago), U.S. intelligence agencies might have relied on “relational databases” from Oracle, IBM (the DB2) or Microsoft (the Sequel server), Jarrell told *GSN,* but such relational databases no longer have the horsepower or speed to keep up with the ever-expanding data-crunching task confronting the intelligence community. “You hit the query button and the answer never comes back,” he said.

Part of the problem, Jarrell explained, is that a relational database requires its computer to compare *every* characteristic of *every* contact at *every* degree of separation against every other characteristic in the entire database. This can take forever.

The “secret sauce” of the technology that underlies Objectivity’s *InfiniteGraph* is that it doesn’t have to compare every characteristic in the database against very other characteristic. Instead, Objectivity has devised a method of attaching unique identifying numbers to each individual in a database, and then attaching the same unique identifying number to those characteristics that are deemed significant and which involve the same person. The judicious use of these unique identifiers enables *InfiniteGraph* to crunch an enormous amount of information -- measured in petabytes and exabytes -- in a matter of seconds, claims Jarrell.

Another difference between relational databases and *InfiniteGraph*, said Jarrell, is that relational databases tend to *centralize* the data-crunching, while his company's approach tends to *distribute* the task of analyzing the information. Centralizing the process tends to slow it down, while distributing the task tends to speed it up and bring it closer to the individual intelligence analyst actually waiting for the results, Jarrell told *GSN*.

Apparently, the U.S. Government has come to recognize Objectivity's contribution to the vital task of connecting the dots. For years, Objectivity's enabling technologies have been used by the CIA and other intelligence agencies to boost the capability of their analysts. While he would neither confirm nor deny that the company's products have been used directly in the crunching of the new data derived from the raid on Osama bin Laden's compound in Abbottabad, Jarrell was proud to describe the role that *InfiniteGraph* will play more generally in a newly-established intelligence center located in northern Virginia.

To understand the likely role of this new Center, let's imagine that among bin Laden's personal property seized by the SEAL team that fateful night in Abbottabad included computers (whose hard drives were still intact), diskettes, USB sticks (which can store mountains of data), diaries, paper files and all sorts of documents. More than likely, this windfall of information was bundled up in Pakistan and transported back to the United States, where it was delivered to the newly-created information processing agency, known as the National Media Exploitation Center, or NMEC. This Center, which has received very little publicity to date, is intended to serve as a central data analysis center for all of the intelligence agencies that are overseen by the Director of National Intelligence, as well as the FBI and other information-gathering government organizations.

In theory, agencies will willingly turn over some of their files, intelligence tips, phone records and other data to the NMEC, so that the organization's technicians and analysts can make sense of the information it receives in a uniform and consistent manner, and then make that information available to any authorized agency that wants to examine the analyzed data. "The Exploitation Center is trying to help the agencies share information," said Jarrell. "We consider it a private social network. It's like the Library of Congress for the intelligence agencies."

It's not surprising that Objectivity should think so highly of the new NMEC. Its enabling technology is playing a key role in boosting the data-processing power of the NMEC's existing databases, and helping to tie those databases together.

"We are the key database backbone for that agency," Jarrell told *GSN*, with an obvious note of pride.