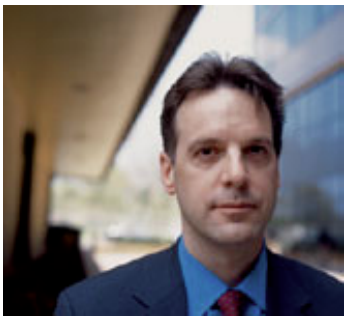
AN MIT ENTERPRISE
TECHNOLOGY
REVIEW
BUSINESS • OPPORTUNITY • IMPACT

[<< Return to article]

Total Information Overload

Co–program manager Robert L. Popp on the U.S. Defense Department's Terrorism Information Awareness project.



Photograph by David Deal.

By Erika Jonietz
Point of Impact
July/August 2003

Robert L. Popp

Position: Deputy director, U.S. Defense Advanced Research Projects Agency Information Awareness Office

Issue: Terrorism Information Awareness. This DARPA project, formerly known as Total Information Awareness, seeks better technologies to detect terrorist attacks but has roused the ire of privacy advocates.

Personal Point of Impact: Co–program manager, Terrorism Information Awareness

Technology Review: There have been wildly varying reports about what Terrorism Information Awareness seeks to do. Some groups opposed to the project have said it includes efforts to link public and private databases, with information ranging from consumer buying habits to medical records, into a giant "metabase." Is there any truth to this?
Robert Popp: First off, let's talk about what Terrorism Information Awareness, or TIA [TEE-ah], is. It's a visionary R&D program that is developing and integrating a variety of information technologies into a prototype system/network to detect and preempt foreign terrorist attacks. As technologists, we are trying to provide the foreign intelligence, counterintelligence, and counterterrorism communities with prototype information technology that will lead to better collaboration, analysis, and decision-making. If we successfully transition these technologies to the operational agencies, we think government decision-makers will be empowered with knowledge about terrorist planning and preparation activities that will help them make informed

decisions to prevent attacks from occurring against the United States.

TR: Will TIA look at U.S. citizens to do any of that?
Popp: No. TIA is not a domestic surveillance capability, nor is any U.S. citizen's privacy changing as a result of TIA. That's one thing that's been widely reported, and nothing could be further from the truth. We are providing operational agencies within the Defense Department and intelligence community with analytical tools that we hope will improve their ability to counter terrorism. These agencies are experimenting with the TIA tools using data and databases they currently have available to them, in accordance with existing laws, regulations, and policies.

TR: So you're not scanning databases.
Popp: Correct. We're not developing technology that will surreptitiously scan or pull data out of a database. TIA is also not creating a grand database of dossiers on U.S. citizens, or developing collections technology to mine transactional or other kinds of data on U.S. citizens that is
prohibited by law.

TR: What kinds of information constitute transactional data?
Popp: Examples might be the purchase of airline tickets to potential attack sites for reconnaissance, the purchase of materials for some kind of bomb, different types of communications transactions—

TR: Like this phone conversation, or an
e-mail?
Popp: Yes. Phone conversations, e-mails, chat messages, newswire stories, et cetera, are all examples of what we consider to be communications transactions.

TR: Where do you think the false perceptions about your work come from?
Popp: Back in November of 2002, as the Homeland Security Bill was being passed, a national newspaper published a column that asserted the bill would permit DARPA to create a system—TIA—to continuously update electronic dossiers on the transactions of every American. As I said earlier, nothing could be further from the truth. But unfortunately a lot of news outlets and Web sites picked up the story and printed this information as if it were fact.

In retrospect, we could have been—and should have been—more outspoken in the public and with Congress about what we were and were not doing in TIA to straighten out the record. The program's name change is designed to help clear up the confusion, to make it absolutely clear that the goals of TIA are to protect the U.S. and its citizens from foreign terrorist threats, period.

TR: So what are the technologies being developed under TIA?
Popp: In broad terms, the technologies we're focused on are collaboration, analysis, and decision support tools; foreign-language translation; pattern recognition and predictive modeling; databases and privacy protection; and biometrics. We have numerous programs developing these technologies, and TIA is the program that integrates these technologies into a unified network/system.

TR: Are any of these technologies anything like those people have expressed concern over—examining public or private-sector databases, for instance?

Popp: Let me answer this by describing the two threads of activity we're pursuing in TIA, namely, an operational thread and a purely R&D thread. This distinction is important to understand because it has been a major source of confusion in the public about what we're doing in TIA.

The premise driving the operational thread is a widely held belief that the data necessary to *effectively* counter the terrorism threat is already in government-owned databases. This was certainly one of the conclusions from the joint House-Senate inquiry of the events that led to the failure of 9/11. For example, two of the 19 hijackers were on the State Department/INS watch list; these same two hijackers were also sought by the CIA and FBI as suspected terrorists. The problem is, the data exist in different databases and are managed by different agencies. Within this thread, TIA is essentially doing two things: first, we're building an R&D network to allow agencies to collaborate and experiment with prototype information technology for counterterrorism, and secondly, we're empowering these analysts with a variety of tools enabling better analysis and decision-making.

There is another community of people who believe that all the data necessary to *effectively* counter the terrorism threat is in fact not entirely in government databases; this premise drives the R&D thread. Instead, there may be more information in the greater information space that might prove valuable for the government to exploit in its counterterrorism operations, but currently this data is not used due to legal or policy restrictions. This thread is testing the hypothesis that when terrorist organizations engage in adverse actions against the United States, they make transactions in support of their plans and activities, and those transactions leave a signature in the information space. Those transactions will most likely span government, private, and public databases.

The challenge for TIA here is twofold: First, is the signature detectable when embedded within a world of information noise? Second, in what part of the information space does that signature manifest itself? Ultimately, our goal within this thread is to understand the level of improvement possible in our counterterrorism capabilities if the government were able to access a greater portion of the information space, while at the same time considering the impact—if any—on the right to privacy, and then mitigate this impact with privacy protection technology. If our research does show a significant improvement in the government's ability to predict and preempt terrorism, then it would be up to the policymakers, Congress, and the public at large—not DARPA—to decide whether to change law and policy to permit access to such data. One footnote: because the government today does not typically access some types of transactional data that *may* prove meaningful, all of this research is being done with synthetic, simulated data.

TR: There's quite a bit of public concern over ideas like that, and about technologies TIA is developing like facial recognition that people worry could be used to track them. How valid are those worries?
Popp: All of us know that technology has shaped how hard or easy it can be to invade people's privacy. The government already has plenty of technology right now that could be used to invade people's privacy if used maliciously. But our laws are what control how the technologies may be used and applied and on what information. Our privacy is protected by the law, not the fact that it is technically difficult for the government to quickly make sense out of the information that it already possesses. But this difficulty sorting through government information does

make it harder to stop terrorists.

Americans' basic privacy will not change unless Congress changes the law on what information the government may collect and how it may be used. The issue in our view really isn't a matter of technology. The difference between the tools and the rules, what can be done and what *may* be done, is crucial. Does protecting privacy mean we should take away the few rudimentary tools that our agencies in the federal government have now to share data and collaborate—phone, fax, and e-mail? Of course not. Ultimately, what TIA is trying to do is give our counterterrorism agencies far better tools to do their job.

Online Extra: Popp on protecting privacy in Total Information Awareness

TR: Privacy advocates and the press have expressed concerns about the kinds of data the Total Information Awareness (TIA) program might be looking at. What sorts of databases would TIA examine?

Popp: TIA, which now stands for Terrorism Information Awareness, is using either foreign intelligence or counter-intelligence data that was obtained legally and may be used by the federal government under existing laws, regulations and policies, or else wholly synthetic, artificial data that we've generated to resemble real-world patterns of behavior. So far our major focus has been on providing entities within the Department of Defense (DoD) and intelligence community with information technologies like collaboration software, analytical tools and decision support aids that they then use in experiments on the data and databases they currently have available to them in accordance with existing laws, regulations and policies. Under no circumstance are we in the TIA program providing any real data or any means to collect real data.

Examples of the types of data being used include foreign intelligence data such as imagery intelligence, signals intelligence, human intelligence; data that is in the public domain such as the World Wide Web and various news feeds such as AP, Reuters and Al Jazeera; and map and geospatial data that is available from a variety of commercial and government sources.

TR: Is there any oversight of the TIA projects, either from within the Department of Defense or from outside?
Popp: Yes, in fact, both. The Secretary of Defense has established an oversight framework that consists of both an internal oversight board and an external Federal Advisory Committee. The charter of the internal oversight board is to oversee and monitor the manner in which TIA tools are being developed and prepared for transition to real world use, as well as establish the supporting policies and procedures as necessary. The board, formed in February 2003, is composed of senior officials within the Defense Department and intelligence community and is chaired by the Undersecretary of Defense for Acquisition, Technology and Logistics. The external Federal Advisory Committee held its first meeting in May 2003, and its broad charter is to advise the Secretary of Defense on the legal and policy issues, particularly those related to privacy, that arise as advanced technologies are being applied in the war on terrorism, such as those being developed by TIA.

TR: Are there any privacy safeguards built into the TIA program?
Popp: Yes—many. In addition to the oversight boards, DARPA has been and continues to be fully committed to managing and overseeing the TIA program in full compliance with the laws and regulations that protect the privacy and civil liberties of all Americans.

As an integral part of our commitment to safeguarding privacy, we are also sponsoring research that aims to create – and assess the merits of – a variety of privacy protection technologies. These technologies will protect not only U.S. citizens' privacy, but also protect the identity of sensitive intelligence sources and their methods which is an analogous problem in our view.

TR: What are those?
Popp: Before describing some of the specific technologies, let me first note that our overarching goal for the privacy protection work we're sponsoring is to enhance both privacy and security. Oftentimes, the debate over privacy and security issues tends to be portrayed as a tradeoff; we don't believe this to be the case. A second important point that is worth reemphasizing: TIA's research and experimentation is using only legally obtained intelligence or counter-intelligence data, or synthetic data. We're exploring a number of privacy protection approaches. Among the more promising techniques are Transformation Spaces, Selective Revelation, Self-Reporting Data, Anonymization, Immutable Audit, and Privacy Appliance. Let me explain these.

Transformation Spaces is an approach that uses well-known concepts of mathematical analysis and encoding techniques to transform data from a plain-text representation to a cipher or mathematical space that is unintelligible to a human. Once transformed, a plethora of data analysis functions or mathematical operations can be applied very efficiently while simultaneously protecting the privacy of the data. TIA hopes to demonstrate that very specific patterns can be developed that describe terrorist signatures. We anticipate these signatures may be buried in an enormous amount of data about everyday worldwide activity that has nothing to do with international terrorism whatsoever. We also think there is a wider range of intelligence data, both classified and open source, which analysts need to search in order to understand terrorist intent. To have any hope of making sense of this, we believe there must be a more structured and automated way of handling this problem. With Transformation Spaces, we're exploring the merits of working within a transformed mathematical space to address this challenge. Because the data is represented in a space that is unintelligible to the human, privacy protection is inherent in this approach.

Selective Revelation would allow incremental access to and analysis of increasingly privacy-sensitive data; analyst knowledge of an individual's identity would occur only after the appropriate legal standards have been met. The approach proceeds incrementally by initially requiring a data owner to release only subsets or statistical interpretations of its privacy-sensitive data to an analyst's query. If the results of the initial query turn out to be meaningful – say the level of suspicion has been heightened – then through appropriate legal frameworks, such as probable cause, more privacy-sensitive data can be released as the analysis progresses. As an example, an analyst can initially issue a pattern-based query indicative of a terrorist attack across a distributed set of data sources to determine if there is any evidence of that pattern. The initial set of results to the query may be so large that it needs to be anonymized or converted into statistics – no identity data is provided. Through iterations, the analyst can sufficiently refine the pattern-based query so that it more strongly implies terrorist activity and only an acceptably few individuals match the query. At that point, the analyst can use query results as evidence to seek legal authorization from an appropriate authority to obtain the identities of the suspicious individuals.

Self-Reporting Data would continuously track and report the location of data and the person accessing it as it transits from one location to the next. We're exploring the efficacy of digital watermarks and similar techniques for this concept.

Anonymization would allow a data owner to release a generalized or obfuscated version of its data to an analyst with a guarantee that the specifics of any privacy-sensitive data in the released data cannot be determined, yet the released data still remains useful from an analytical point of view. Examples of identifiers that typically would be anonymized include name, address and telephone number.

Immutable Audit would automatically record all accesses to data immediately and permanently, with no possibility that audit records can be altered or tampered with. Moreover, to prevent against potential abuses by malicious insiders or agents, the immutable audit will be designed to detect with high probability any misdeeds, and encrypt and transmit the audit records to an appropriate trusted third party oversight authority. We will also develop tools to query and analyze audit data in on-line real-time scenarios as well as in off-line batch mode. The contents of the audit log may contain fields such as the identity of the government user; the authorizations being used; the date and time of the entry; the data requested; and the data returned.

Privacy Appliance is a novel concept that would employ a separate hardware device placed on top of a database, metaphorically of course. It would serve as a trusted, guarded interface between the user and the database –analogous to a firewall – and would implement several privacy functions and mechanisms to control and enforce access rules and accounting policies. It would also explicitly publish the details of its operation; it would be tamper-resistant and cryptographically protected; it would enforce business rules established between the database owner and the user issuing a subject- or pattern-based query; it would verify the user's access permissions and provide access only to authorized users; it would filter queries to permit only those that do not violate privacy; it would verify the credentials of the user issuing the query that are packaged with the query with respect to specific legal and policy authorities under which the query has been conducted; and it would initiate an immutable audit log capturing the user's activity and transmitting it to an appropriate trusted third party oversight authority to ensure abuses are detected.

TR: In February, President Bush signed a spending bill that included an amendment introduced by Sen. Ron Wyden (D-Ore.). The Wyden amendment called for a report on TIA to Congress. Has the amendment limited the project in any way so far?
Popp: First, I'm happy to report that the congressionally mandated report on TIA was delivered to Congress on May 20, as the Wyden amendment had ordered. Now, with respect to the limitations on the TIA project as a result of the Wyden amendment, it does pose limits on the deployment and implementation of TIA technology (and technology from related programs within the Information Awareness Office) to any department, agency, or element of the Federal Government that is not engaged in lawful military operations of the United States conducted outside the United States or lawful foreign intelligence activities conducted wholly against non-US persons. We have and will continue to comply with these limitations.

Erika Jonietz is a senior associate editor at *Technology Review*.