the Atlantic

The Military Tackles Anonymous Data-Sharing

DARPA is tackling online privacy. But can you trust them?

Patrick Tucker

March 13, 2015

The average, technologically connected American worker produces some 5,000 megabytes of digital data a day, enough to fill nine CD-ROMs. Only a small fraction of it is stored permanently or is clearly related to a specific identity, but the sheer volume of digital exhaust that is daily life has turned privacy into an endangered entity – and a growing national security concern.

On Wednesday, the military's Defense Advanced Projects Research Agency, or DARPA, put out an agency announcement on a program that seeks to restore some semblance of privacy to the online world. The so-called **Brandeis** program, named after the late U.S. Supreme Court associate justice and privacy advocate Louis Brandeis, seeks to build "information systems that can ensure private data can only be used for its intended purpose and no other."

Privacy deprivation may be a fact of modern, inter-connected life, but if passwords, files, and personal location data can't be protected, then neither can vital pieces of information. That vulnerability can contribute to industrial theft and sabotage and worse, and the recent **Sony hack** and **CENTCOM Twitter snafu** foreshadowed this naked future. It's the military's job to protect the country from national security threats. Question is: in the wake of the Edward Snowden scandal and what it revealed about NSA data collection practices and capabilities, how much does anyone trust the military to, essentially, build them a privacy machine?

The specific type of data that the DARPA program seeks to protect is your transactional data or data that you knowingly stream to a site or party. But it comes with a crucial caveat – the Brandeis program addresses data "that is knowingly provided to a third party, as opposed to data collected as a byproduct of interacting with the network or a system," according to the DARPA announcement.

That caveat is important for two reasons: First, the goal is not just the protection of privacy but also the protection of privacy while users are *engaged in the act of sharing data*. Second, by focusing on information that citizens knowingly give to third parties rather than inadvertently provide as a result of merely interacting with machines the DARPA program rules out building any future information system that might, somehow, get in the way of the military or law

enforcement collecting signals intelligence as part of investigations (to the continued objection of privacy advocates).

Why is protecting data *sharing* important? On this count, the DARPA announcement is so strident on the social value of **yottabytes** of user-generated data that the casual reader might think the notice came from the Google press office.

Data sharing will create "personal medicine (leveraging cross-linked genotype/phenotype data), effective smart cities (where buildings, energy use, and traffic controls are all optimized minute by minute), detailed global data (where every car is gathering data on the environment, weather, emergency situations, etc.), and fine grained internet awareness (where every company and device shares network and cyber-attack data)."

Without strong privacy controls, none of those futuristic visions can be realized. But protecting the privacy of people who are voluntarily sharing data is no straight-forward undertaking. The same correlational analysis that can reveal a relationship between a certain protein construction and cancer can reveal the individual that volunteered that genomic information. A person's electricity-usage patterns are distinct, based on an individual's schedule, devices, habits, etc. All of that speaks to identity. For that reason, the idea of rendering data fully and permanently anonymous is a dubious one among much of the privacy community. The DARPA researchers acknowledge that obstacle by pointing to the **work** of Carnegie Melon University's Latanya Sweeney, who has shown that gender and zip code are enough to identify 87 percent of individuals by name.

But just because data is revelatory doesn't mean that all data shared with a third party is viewable everywhere or at low cost. Ideally, the user who is sharing the data should be able to control how it's viewed, rather than that responsibility falling on the third party. The point of the project is ensuring that those telltale data bits don't wind up in the wrong places.

The program, which will go on for four years, will break participants up into three technical areas. The first area is privacy computing. It will take recent strategies and innovations in privacy protection, such as secure database querying, multiparty computation, (a subset of encryption that allows computers to share data even if they don't fully "trust" one another), and **remote attestation**, (which allows a computer server to verify and authenticate the configuration of computer attempting to make contact), and use them to build a more holistic privacy computing framework. Research into these methods has "been promising, but to date these techniques suffer from significant practical limitations in flexibility, scalability and performance," DARPA writes in its announcement.

The second area will focus on human, digital interaction. The goal here is to create systems that can essentially predict the difference between shareable data and more private data, help users understand the privacy tradeoff of different sharing decisions and give users faster and firmer control of what data they share and with whom they share it.

The third area looks to build experimental systems to "provide the platforms on which to test these ideas in practice" — in essence, to create a privacy machine.

Jeremy Gillula, a staff technologist with the privacy watchdog group the Electronic Frontier Foundation, said he had no problem trusting the U.S. government to create privacy solutions, "as long as the research is published and the information is shared with the greater privacy-technology community." He characterized the challenge that DARPA was undertaking in with Brandeis project as an ambitious one.

"The research community has made some significant progress (differential privacy, secure multiparty computation, etc.), but all of these methods have some sort of weakness," Gillula said. "With that said, focusing on ways to scale these methods would certainly be a step in the right direction. It certainly wouldn't solve the whole privacy problem, but at the same time solving the whole privacy problem would be way too big to tackle in a single project."

Julian Sanchez, a senior fellow with the libertarian-leaning Cato Institute, told *Defense One* he found it "somewhat reassuring to see the DARPA solicitation indicates a preference for open source proposals, since that would in practice be a prerequisite for any tools that might be made available for broader public use."

He added that it was, "certainly valuable for DARPA to be sponsoring research into protective information sharing systems of this kind — and indeed, this should have been a priority long ago."

"But citizens and courts shouldn't let themselves be gulled into blessing government data dragnets on the premise that fancy technology will somehow guarantee the data is only ever used by good people for good purposes."

Gillula, sounded a similar note of enthusiasm for the project, but skepticism about a total privacy solution coming out of the government — or anywhere else.

"The program is focused on data willingly shared, and won't help with non-consensual third-party tracking/collection, which is a bigger concern (and more problematic) in general," he said. "The program is fine for what it's doing, but it's definitely not going to solve the greater privacy problem we face as a society."

In other words, the military's privacy machine may facilitate safer information sharing, but fixing privacy — and mistrust of government on the issue — will take more than a press of a button.