usacm ACM US Public Policy Council

August 1, 2013

Mr. David Medine
Chair
Privacy and Civil Liberties Oversight Board
2100 K Street Northwest
Washington, D.C. 20427

Dear Chair Medine:

We are writing to submit comments in connection with the July 9, 2013 workshop of the Privacy and Civil Liberties Oversight Board. We are the U.S. Public Policy Council of ACM (USACM), a community of technical experts representing ACM — the Association for Computing Machinery — a major technical and professional society with members involved in all aspects of computing and information technology. ACM's members have decades of experience in the development, implementation and use of databases, with consideration for ensuring the privacy and security of the information in those databases. As I indicated to you in our phone conversation, USACM stands ready to provide assistance, as you request it.

Based on the discussions at the Board's July 9, 2013 workshop on the National Security Agency surveillance programs authorized by the USA Patriot Act and the Foreign Intelligence Surveillance Act, we offer the following comments. Our intention is to address technical questions and assumptions connected to these surveillance programs and their implications for privacy and security. We are specifically not addressing the policy issues about the scope and legal basis for these systems. Given the classified nature of these programs, and the incomplete information publicly available, our comments are necessarily high-level in nature. Should you wish further explanation or have follow-up questions, please do not hesitate to contact our Public Policy Office.

Limitations of Computing

Before discussing technical issues specific to the surveillance programs being considered by the Board, it is worth taking time to note the limitations of computing and computing systems. Policymakers often presume that computer systems can do more to achieve a desired policy objective than is actually the case. Part of USACM's work is dedicated to making sure policymakers understand what computing can and cannot do in support of policy objectives.

Much like people, computers are fallible. They can break down or produce erroneous results – especially if they are not maintained properly. Software can have bugs, and computer systems have many threats to contend with both from inside and outside of the organizations where they are used. These threats include, but are not limited to: viruses, botnets, other malware, poorly patched software, faulty hardware, improperly implemented security measures, and users who fail to maintain their systems. One of the more pernicious issues is that of an "insider threat" –

someone within an organization who takes advantage of lax security and other procedures to release a great deal of protected information. Alternatively, someone with great authority may be able to insist that access controls be overridden and audit results be ignored or made unavailable for review without the proper policies in place to prevent this.

Even when a set of desired controls has been determined, implementing those controls may be difficult in every place where the data resides. In a few cases (reportedly for some NSA surveillance data), highly sensitive data is kept segregated. But in other cases, the data flows under analysis are mixed and diffused until they are no longer separable. In addition, various collected items are stored in data formats used by specialized systems in a number of different places, which results in poor data coherence. Few organizations are well-positioned to track all of their data flows. Even if they can, it is difficult to transform high-level policies on data controls into appropriate rules that are executable in each specialized system.

Databases are safe and reliable only if the information they contain is accurate. A database system cannot, by itself, determine if collected information is accurate or reliable. Bad or incomplete data will lead to flawed searches and results, which will contribute to flawed decisions. For instance, an electronic employment eligibility verification system will be able to confirm that people with certain credentials can legally work in the United States. But if those credentials are flawed or incomplete, the system's ability to verify employment eligibility suffers.[1]

Computing can allow for the collection, analysis and exposure of information in volumes and ways that were not previously possible (or imaginable). But computing alone cannot provide expert judgment on the meaning, utility, and uses of that information. Without taking the proper steps to implement and execute policies for controlling the collection, security, integrity, audit, accuracy, and use of information, application of computing tools may lead to policy problems that can be harder to solve than the problems those tools were intended to solve.

We suggest that it would be valuable to conduct a systems engineering analysis of the data collection and analysis structure(s) developed – intentionally and unintentionally – by the NSA programs. This would include the specific technical requirements of these systems, the operational assumptions underpinning the programs, and the practices involved. From there it should be simpler to observe and analyze the complex trade-offs involved among national security interests, technical capabilities, privacy and civil liberties principles, and other factors of concern.

Unfortunately, the limited access to information and its classified nature make it difficult for outsiders to effectively analyze the needs of, and assumptions made by, all parties in this discussion. At the least, we suggest that such an analysis should be conducted within the

[1] A few of our members have provided Congressional testimony on this topic in recent years, and we would be happy to share this if you would like to see it.

government, and reviewed by cleared legal and technical personnel outside the agencies involved. Critical assessment of the programs, including whether or not less invasive methods and practices could achieve the same results, can help address the concerns of many parties.

Technical Issues

Identification of persons subject to surveillance

The programs under consideration require that collected information can only be returned for specific persons in certain circumstances.[2] The ability to check these circumstances may depend on several distinct matters: locating a communications device and linking that device to a particular individual or individuals, and knowledge about the individual's status. Each of these matters has considerable uncertainty, which technology will never completely remove.

Technologies exist that can determine the location of particular communications devices, but there are also means of intentionally circumventing those technologies or otherwise concealing the location of a particular individual and/or device for various reasons whether they be for privacy-enhancing or malicious purposes.

Network analysis, which often relies on the use of metadata, can be very effective in finding nodes that see a great deal of activity. Recent research has suggested a relatively low degree of separation (the number of hops before any person is connected to all other people in the network) between people in social network sites, but other networks may have higher degrees of separation. Part of the search criteria should include an assessment of the minimum degree of separation needed to conduct an effective search to avoid false positives, such as those caused by spoofed IP addresses or botnet activity.

Policies that assume collected data is perfect ignore critical issues that affect the desired outcomes of searching that data. Policies need to deal explicitly with uncertainty, expressing confidence in particular assertions. Implementation of these policies will address specific technologies, and those measures need to be reviewed by privacy officers.[3, 4]

[2] Meaning that the search software will examine other records, but only return the information that fits within the authorized parameters.

[3] Assertions might, for example, be about location, or who owns a device, whether that owner is a non-US person, or whether the owner is linked to terrorism. Each of these assertions is probabilistic. Debate about the appropriate levels of protection for U.S. persons and non-U.S. persons is outside our technical expertise, and is not addressed here.

[4] The strength of the links between people is another assertion, and not all links are equal in meaning. For instance, both parties communicating with the same suspected terrorist is a much different link than both parties communicating with a major airline.

Data minimization

The breadth of data collection raises concerns about the ability to effectively secure this data and maintain the privacy of persons whose information is collected. If the NSA is to collect and store data within its own infrastructure, we would encourage it to follow general fair information practices of minimization. This includes both *data minimization* – collecting only the information needed for a particular purpose, and *use limitation* – avoiding the use of information for additional purposes not connected to the ones for which data was initially collected. This would reduce the amount of information collected, and the associated demands on personnel and infrastructure to effectively manage its storage, analysis and access. Careful consideration should also be given to expiration of data: data becomes less reliable as it ages, while continuing to tempt parties to use it for other purposes.

A massive dataset is a large target, especially if the dataset is stored or accessible via Internet-connected systems. To the extent practical, data searched should be kept where it originated (e.g., at the communication providers' data centers), and both searches and analyses distributed across those locations. This has already been shown to be feasible for certain statistical computations in health care.

But such distributed work is not easy. Conducting searches and analyses across different storage locations requires that all sources respond promptly to queries, that queries are adapted so they do not reveal the query-submitter's interest to data holders or those eavesdropping on them, that analyses are amenable to being done in pieces and assembled later, and that rich connections between classified and critical commercial systems are tightly maintained without reducing their security.

Data access controls

Board members expressed an interest during the workshop in finding technical means to limit the uses of collected data. We agree with Steve Bellovin who indicated during the July 9 workshop that limiting access is primarily a policy question rather than a problem in search of a technical solution. Policies can be implemented by technical means, but choosing particular technologies cannot guarantee limitations on use of collected information.

Because these programs collect a great deal of data that may be outside the scope of specific, authorized searches, the temptation for decision makers to expand the scope of what is authorized is perhaps greater than in other circumstances. There will also be the temptation, if only for expediency, for operational personnel to take advantage of gaps in enforcement. Only consistent implementation of access controls, with regular audits and timely, forceful corrective actions can ensure that inappropriate use (as well as outsider access) of collected data is minimized (not eliminated).

Technical assistance for FISA courts

Many computing professionals from outside of government have experience in classified environments, and would be able to assist the FISA courts in their reviews and deliberations in answering technical questions about the programs and cases the courts review. The courts may wish to devise an ongoing method of obtaining needed technical advice from computing professionals who possess the necessary clearances. This takes on more urgency, as the FISA court does not benefit from the usual adversarial discovery and argument present in other courts.

We appreciate the opportunity to provide comment to the Privacy and Civil Liberties Oversight Board. Should you have any questions on the above, or need additional information, please contact our Public Policy Office at acmpo@hq.acm.org, or our Director of Public Policy, Cameron Wilson, at 212-626-0541.

Regards,

Eugene H. Spafford, Ph.D.
Chair, U.S. Public Policy Council
Association for Computing Machinery