Government Eavesdropping—How They Do It

By Louis P. Solomon

Last month I wrote a column expressing my concern about the legal issues associated with the separation of powers within the federal government as it relates to electronic eavesdropping. One of the key technical issues is the following: How do they electronically eavesdrop? What is the technology that allows the consideration and monitoring of billions (!) of communications a day?

The over-riding concept is that there *are* billions of communications *per day*. Google alone has approximately 400 million queries daily. And those are just queries to one web site; they are not email messages. It is estimated that there are 1 billion email messages a day, sent and received world-wide. And the number of telephone calls, with the advent of cellular technology, is estimated to be much larger. It is clear to even the most casual observer that technology must play a major role in the entire concept of practical eavesdropping: obtaining information of value. The amount of communication is simply too large to scan by people, and is growing each year. So, what methods are used to look at the large number of enormous data bases?

The method used to scan data bases is known as Data Mining. Data Mining uses many different techniques and is applicable to all data bases, but in particular works most efficiently on computer stored data bases. Let me focus on just one technique that is employed in the searching of data bases: *Keywords*. There are many others.

As most of you know, keywords is the method by which drives the operation of the search engines in Google. If you enter Google and search for the existence of a generic book about explosives (say), then you might enter the keywords: explosives, bombs, and detonators. Google search engines look for all the records in its enormous data base that contains those keywords. A data record for a particular book, that the writer or publisher describes using the keywords explosives, bombs, or detonators, would be found by the Google search engine. This book would be presented to the shopper for possible selection and purchase. In such large data bases there could be hundreds of books presented to the potential buyer. Now, let's make the investigation process a little more complex. Let us assume that you (as a government agent) wanted to find the names of all people who had asked about the keywords explosives, bombs, and detonators. This list is easily constructed by the computer. If you were only interested in people who had searched for books using the keywords explosives, bombs, and detonators, then your search would be over. But, what if you had another search you wished to implement? In fact, let us assume that you decided to search for the people who were also interested in the geography of the Middle East. A search based upon appropriate keywords for Middle East geography would also yield a list of people.

If you cross check the two lists, i.e, compare the list of the people who searched for books using the keywords explosives, bombs, and detonators with the list of the people who were interested in the geography of the Middle East there could be some people who were on both lists. Does that mean that they are terrorists? Of course not. But, this is the general method that is used. The searches are for *patterns of interest and behavior* that might yield a list of people who exhibit the patterns of interest and behavior that the government (or other organizations) thinks *might* indicate possible terrorists (or people of interest). These searches are performed on a large number of data bases.

The interesting point here is the assumption that people's actions reflect their interests (not too hard to draw that inference) and if you wish to find people who have interests that you consider would make them worthy of further investigation, then you have to have access to large numbers of enormous data bases which you can then subject to data mining. Google is an owner of one such data base. The telephone, telegraph and cable companies also own such data bases. It should also be clear that if you are going to search such data bases it can only be done by large, very fast computers. It is not possible for human beings

to review the mass of information that is passing through the world information channels.

Most data mining activity yields information that is unusual, and frequently is of great interest to business organizations. It is not unusual for data mining to find correlations between habits that ostensibly are completely unrelated to one another. One author cites a case where a major grocery chain found a bizarre correlation: for some reason, people who bought large quantities of beer also bought large amounts of diapers. There is no reason advanced for this correlation; but the grocery chain did put the diapers near the beer, and their profits rose measurably.

The rapid growth of computers, their extraordinary data handling capability for virtually no cost has led to implementation of data mining methods. The people who employ such methods are well aware that there are substantial pitfalls in finding correlations that are discovered simply through the process of turning on the computers and let them solve all the problems. Some times there occur correlations that, when considered with a skeptical eye, are simply events that randomly occur in statistical studies. But, some times these events are not random, and lead to more substantial inquiries.

I personally think that the real use of data mining, in particular in the application of searching for potential terrorists and other enemies of the United States, is to recognize a “person of interest” through their actions. Once the person is identified then actual investigation, by human beings with human intellect, can begin. A real concern within American society is the loss of privacy and the possible incorrect and inappropriate inferences that may be drawn about some innocent member of society. In the current use of data mining applied to looking for possible terrorists, many perfectly innocent Americans have been considered, and found to be just what they appear to be: innocent Americans with unusual interests. A concern that always arises in the application of extensive data mining of all Americans is the possible loss of our privacy. In practice, just living in our current world has caused us to lose our privacy in ways that most of us simply do not recognize. We believe that our social security, driver’s license, and bank numbers are all carefully held and are private. But a few moments of earnest search on the Internet will disabuse us of that notion. All the private data of individuals is there for people to find, study, and possibly use for nefarious purposes.

The application of the concepts of data mining, coupled with methodology that can only be implemented using large very fast computers scanning many enormous data bases is very valuable in our search to support our national security interests. Unfortunately, these same capabilities might threaten the privacies in our personal lives that we, as Americans, hold so dear. The federal government must design and implement safeguards that make use of modern technology and protect the privacy of individual citizens.