GlobalResearch
Center for Research on Globalization

Pentagon sets its sights on social networking websites

By Paul Marks

Global Research, June 13, 2006
New Scientist magazine, 09 June 2006, page 30 9 June 2006

Theme: Police State & Civil Rights

From issue 2555 of New Scientist magazine, 09 June 2006, page 30

"I AM continually shocked and appalled at the details people voluntarily post online about themselves." So says Jon Callas, chief security officer at PGP, a Silicon Valley-based maker of encryption software. He is far from alone in noticing that fast-growing social networking websites such as MySpace and Friendster are a snoop's dream.

New Scientist has discovered that Pentagon's National Security Agency, which specialises in eavesdropping and code-breaking, is funding research into the mass harvesting of the information that people post about themselves on social networks. And it could harness advances in internet technology – specifically the forthcoming "semantic web" championed by the web standards organisation W3C – to combine data from social networking websites with details such as banking, retail and property records, allowing the NSA to build extensive, all-embracing personal profiles of individuals.

Americans are still reeling from last month's revelations that the NSA has been logging phone calls since the terrorist attacks of 11 September 2001. The Congressional Research Service, which advises the US legislature, says phone companies that surrendered call records may have acted illegally. However, the White House insists that the terrorist threat makes existing wire-tapping legislation out of date and is urging Congress not to investigate the NSA's action.

Meanwhile, the NSA is pursuing its plans to tap the web, since phone logs have limited scope. They can only be used to build a very basic picture of someone's contact network, a process sometimes called "connecting the dots". Clusters of people in highly connected groups become apparent, as do people with few connections who appear to be the intermediaries between such groups. The idea is to see by how many links or "degrees" separate people from, say, a member of a blacklisted organisation.

By adding online social networking data to its phone analyses, the NSA could connect people at deeper levels, through shared activities, such as taking flying lessons. Typically, online social networking sites ask members to enter details of their immediate and extended circles of friends, whose blogs they might follow. People often list other facets of their personality including political, sexual, entertainment, media and sporting preferences too. Some go much further, and a few have lost their jobs by publicly describing drinking and drug-taking exploits. Young people have even been barred from the orthodox religious colleges that they are enrolled in for revealing online that they are gay.

“You should always assume anything you write online is stapled to your resumé. People don’t realise you get Googled just to get a job interview these days,” says Callas.

Other data the NSA could combine with social networking details includes information on purchases, where we go (available from cellphone records, which cite the base station a call came from) and what major financial transactions we make, such as buying a house.

Right now this is difficult to do because today’s web is stuffed with data in incompatible formats. Enter the semantic web, which aims to iron out these incompatibilities over the next few years via a common data structure called the Resource Description Framework (RDF). W3C hopes that one day every website will use RDF to give each type of data a unique, predefined, unambiguous tag.

“RDF turns the web into a kind of universal spreadsheet that is readable by computers as well as people,” says David de Roure at the University of Southampton in the UK, who is an adviser to W3C. “It means that you will be able to ask a website questions you couldn’t ask before, or perform calculations on the data it contains.” In a health record, for instance, a heart attack will have the same semantic tag as its more technical description, a myocardial infarction. Previously, they would have looked like separate medical conditions. Each piece of numerical data, such as the rate of inflation or the number of people killed on the roads, will also get a tag.

The advantages for scientists, for instance, could be huge: they will have unprecedented access to each other’s experimental datasets and will be able to perform their own analyses on them. Searching for products such as holidays will become easier as price and availability dates will have smart tags, allowing powerful searches across hundreds of sites.

On the downside, this ease of use will also make prying into people’s lives a breeze. No plan to mine social networks via the semantic web has been announced by the NSA, but its interest in the technology is evident in a funding footnote to a research paper delivered at the W3C’s WWW2006 conference in Edinburgh, UK, in late May.

That paper, entitled *Semantic Analytics on Social Networks*, by a research team led by Amit Sheth of the University of Georgia in Athens and Anupam Joshi of the University of Maryland in Baltimore reveals how data from online social networks and other databases can be combined to uncover facts about people. The footnote said the work was part-funded by an organisation called ARDA.

What is ARDA? It stands for Advanced Research Development Activity. According to a report entitled *Data Mining and Homeland Security*, published by the Congressional Research Service in January, ARDA’s role is to spend NSA money on research that can “solve some of the most critical problems facing the US intelligence community”. Chief among ARDA’s aims is to make sense of the massive amounts of data the NSA collects – some of its sources grow by around 4 million gigabytes a month.

The ever-growing online social networks are part of the flood of internet information that could be mined: some of the top sites like MySpace now have more than 80 million members (see Graph).

The research ARDA funded was designed to see if the semantic web could be easily used to connect people. The research team chose to address a subject close to their academic

hearts: detecting conflicts of interest in scientific peer review. Friends cannot peer review each other's research papers, nor can people who have previously co-authored work together.

So the team developed software that combined data from the RDF tags of online social network Friend of a Friend (www.foaf-project.org), where people simply outline who is in their circle of friends, and a semantically tagged commercial bibliographic database called DBLP, which lists the authors of computer science papers.

Joshi says their system found conflicts between potential reviewers and authors pitching papers for an internet conference. "It certainly made relationship finding between people much easier," Joshi says. "It picked up softer [non-obvious] conflicts we would not have seen before."

The technology will work in exactly the same way for intelligence and national security agencies and for financial dealings, such as detecting insider trading, the authors say. Linking "who knows who" with purchasing or bank records could highlight groups of terrorists, money launderers or blacklisted groups, says Sheth.

The NSA recently changed ARDA's name to the Disruptive Technology Office. The DTO's interest in online social network analysis echoes the Pentagon's controversial post 9/11 Total Information Awareness (TIA) initiative. That programme, designed to collect, track and analyse online data trails, was suspended after a public furore over privacy in 2002. But elements of the TIA were incorporated into the Pentagon's classified programme in the September 2003 Defense Appropriations Act.

Privacy groups worry that "automated intelligence profiling" could sully people's reputations or even lead to miscarriages of justice – especially since the data from social networking sites may often be inaccurate, untrue or incomplete, De Roure warns.

But Tim Finin, a colleague of Joshi's, thinks the spread of such technology is unstoppable. "Information is getting easier to merge, fuse and draw inferences from. There is money to be made and control to be gained in doing so. And I don't see much that will stop it," he says.

Callas thinks people have to wise up to how much information about themselves they should divulge on public websites. It may sound obvious, he says, but being discreet is a big part of maintaining privacy. Time, perhaps, to hit the delete button.

The original source of this article is New Scientist magazine, 09 June 2006, page 30

Articles by: **Paul Marks**