Abstract

Corporations are facing increasing demands to monitor their compliance with security and privacy policies and regulations. There are a number of activities that take place within corporations that are both related to security and privacy, and amendable to social network analysis. The current study focused on three such activities: password sharing, transmission of confidential information, and the sharing of salary data. Social network analysis of communications patterns within an organization was able to identify key creators and sharers of passwords, and an analysis of the passwords themselves showed that quality was clearly a problem. This suggests that password management practices were a problem for this organization. Analysis of confidential information flows revealed two highly interconnected groups sharing information freely among themselves. It appears that confidential information was poorly handled within this organization. Salary data, on the other hand, was only shared among small, isolated groups, and this represents an effective and appropriate information privacy practice. In each of these cases, the value of social network analysis was demonstrated and suggestions were made for developing automated compliance tools for security and privacy. Current work is focused on creating models of information flow within groups to detect inappropriate activities.
There has been a lot of discussion this week about “dark networks” – primarily those associated with terrorism. I want to talk about a different kind of dark network – a group of people acting somewhat badly, perhaps without even knowing it.

The “Little Rascals” were a group of kids who often behaved badly, without realizing it, within a positive social context.
We are talking about dark networks working inside a bright network. People behaving badly within a context of good behaviors.
Data breaches are becoming far too common. Some are caused by laptop theft, hacks, and sloppy data handling procedures.
Some data breaches are caused by inappropriate insider activities.
Numerous laws requiring disclosure of data breaches. Here are some examples from the US.
Common solutions to preventing privacy breaches involve searches, traffic monitoring, or encryption.

Our approach has been to augment these with Social Network Analysis. We are developing methods to discover networks of inappropriate activity, based on a variety of search and monitoring activities.
In order to test our methods and provide a publically-visible example of our analysis, we have been looking at the well-know Enron email corpus. We have been looking in the Enron corpus for examples of interesting social behavior related to security and privacy.

Enron filed for bankruptcy protection in the Southern District of New York in late 2001 and selected Weil, Gotshal & Manges as their bankruptcy counsel. Enron employed around 21,000 people (McLean & Elkind, 2003) and was one of the world’s leading electricity, natural gas, pulp and paper, and communications companies, with claimed revenues of $111 billion in 2000. Fortune named Enron "America's Most Innovative Company" for six consecutive years. It achieved infamy at the end of 2001, when it was revealed that its reported financial condition was sustained mostly by institutionalized, systematic, and creatively planned accounting fraud. Enron has since become a popular symbol of willful corporate fraud and corruption.

We developed our own methods to clean and search the data.

Data comes from: http://www.cs.cmu.edu/~enron/

Method

- 517,431 email messages
- headers parsed for From, To, Date …
- alias substitution
- clean duplicates left 250,641 unique messages from 86,199 email addresses
- then search messages based on criteria of interest: security and privacy-related language
- forensic & remediation analysis based on communication patterns
This is the data I presented at last year’s Sunbelt conference. We were able to uncover interesting examples of password sharing, and key nodes responsible for issuing passwords (and doing it badly). A paper based on this data is available on my web site.
Here is new data looking at people discussing the salary of Enron employees. We found this by looking for salary-related language in the body of the email messages.

Blue nodes are external to Enron.

Very sparse network (average density 1.85%) with 12 components. Thus, salary information not shared widely.

* most messages sent between managers to discuss offers to job candidates or changes to employee salaries

* one example of a headhunting firm presenting potential job candidates

• two examples of employees being recruited by other companies
Here is the data when we look for the word “confidential” in the Subject header. This is often used to flag correspondence to/from lawyers, and is an attempt to protect the communication with a form of lawyer-client privilege.

This is a large network with 850 nodes. Clearly, lots of people felt they were in need of legal protection.

Blue nodes are external to Enron

850 nodes, 1579 ties

Density = 0.26%

average degree 2.4 (in or out)
This is the core of the confidential network.
Method 2: k-core > 4, size based on out-degree

Key Nodes:
1 Government Affairs exec., involved with pricing schemes
6 Government Affairs exec., involved with influencing government probe
8 Government Affairs exec.
16 in-house lawyer, “knows where the bodies are buried”
29 Government Affairs exec., knew of pricing schemes

External nodes are at two law firms.

There is a secondary cluster of people trying to protect their correspondence. They have not been identified in any of the court documents as key players, and yet something seems to be going on. If this was an active investigation, they might be worth investigating.
Searched for “Raptor” in the Subject header.

The “Raptors” were “Special Purpose Entities” setup by Enron to hide losses.

The “Raptors” were found to be illegal and they were basis for many of the criminal convictions.

This is a fairly small network with a high number of connections.
This is the core Raptors network. Method was K-cores, k > 4

Nodes 12 and 16 were responsible for Enron Finances and were eventually convicted

But there were lots of other people who seemed to be key players in the conversations:

2 ???
6 ???
8 ???
22 in-house lawyer
26 in-house lawyer

Clearly, some people would be worth talking to.
Messages from/to 23 people eventually convicted of crimes (although at least one under appeal)

Nodes in blue are the criminals.

5363 nodes, 8775 ties

The “criminals” were deeply tied to the rest of the organization.
k-cores $k > 10$

blue nodes were convicted

This is what a conspiracy looks like. A collection of people acting together. It is interesting to note that there does not seem to be other people at the core of this network.
• detecting networks of people behaving badly together
• uncovering possible causes of security and privacy breaches
• detecting persons of interest during investigations
• scoping inappropriate activity through ‘naïve moles’
• suggesting “direct and indirect” remedial measures
• analysis of specific ‘emotional’ networks to uncover inappropriate or covered-up activities (regret analysis)

• combining network and workflow analyses (anomaly detection)
http://www.andrewpatrick.ca