https://doi.org/10.48009/1_iis_2010_352-357

DATA MINING AND SOCIAL NETWORKING SITES: PROTECTING BUSINESS INFRASTRUCTURE AND BEYOND

Richard Bassett, Western Connecticut State University, bassettr@wcsu.edu
Tiffany Chamberlain, Western Connecticut State University, tmchamberlain82@hotmail.com
Stephen Cunningham, Western Connecticut State University, cunningsteve@gmail.com
Gregor Vidmar, Western Connecticut State University, g.vidmar@yahoo.com

==========

ABSTRACT

Social networks are not new to the IT landscape, having morphed from bulletin boards and chat rooms to the desktop and mobile application tools such as Facebook, Twitter and LinkedIn which millions of people use on a daily basis. These free and easy to use cloud based applications provide users and companies a continuous stream of valuable communication, offering businesses a broader reach and frequency of their branding messages. Unfortunately, in concert with businesses and private users, hackers have joined in seeking to fulfill their own malicious agenda by exploiting information posts in social networks.

INTRODUCTION

The exponential growth of social networks has given rise to great concerns over personal privacy [3] and the many reasons why it is critical to limit the disclosure of personal data on social networks [9, 11].

While considerable volumes of literature focus on the potential personal perils of data disclosure of social networks, very little attention has been has been paid to the organizational risks which businesses face due to the unintended disclosure of company data by its employees.

Employees who have participated in some form of security awareness training are less likely to intentionally divulge potentially damaging information about their company or colleagues to outsiders [5, 17]. But while using social networks such as Facebook, Twitter and LinkedIn users are often unaware that the small bits of data that they are posting about their daily activities can be data mined and reassembled by semi-skilled outsiders, giving an unintended risky inside view of a company.

However, even though this risky environment has evolved, minimizing these risks can be accomplished through the deployment of defensive strategies that include awareness, monitoring and logging, acceptable use policy and enforcement combined with some commercially available technologies.

ENTERPRISE VULNERABILITIES

The usual suspects of vulnerability in a business such as "Social Engineering" [16], "Drive-by Downloads" [14], "Hi-jacked Sites"[12] and others are commonly present in the social networking landscape, opening companies up to viral attacks from within. As users are allowed to post information to these forums, they may also release confidential information about clients, patients or the firm itself. Whether intentional or not, regulatory compliance to the Health Insurance Portability and Accountability Act (HIPAA), Sarbanes Oxley Act (SOX), Gramm Leach Bliley (GLB) or Payment Card Industry (PCI) place regulatory compliance requirements on organizations to protect trusted information.

David Hricik, Professor of law at the Mercer University School of Law brings attention to attorneys participating in social network services, illustrating that conversing with persons online may infer a client-lawyer relationship and the privacy rights assumed may be enforced. Online postings could disqualify a firm from representing other clients since they now have confidential information. [10]

In the medical field clinics and organizations supporting patients with HIV status, pregnancy termination, and history of mental health problems or drug and alcohol abuse can be the source for their privacy being compromised. [1]

For corporate organizations, social networks can also serve as an avenue for insider leaks, which, according

to a 2007 CSI/FBI survey, are the most prevalent security threat for organizations. [4]

Apple recently posted a notice seeking antenna engineers on the heels of the apparent iPhone 4G difficulties. News sources considered this a confirmation of hardware issues. These are but a few of the scenarios to be wary of when a business turns to networking sites. Companies need to treat public postings with caution, as if they were posting their personal vacation schedule on Facebook. [15]

DATA MINING

Data Mining can be imagined as taking many individual pieces of data and combining them in ways which create a larger holistic informational picture by giving users the ability to analyze large numbers of seemingly unrelated elements of data to assist in discovering trends and patterns. [8]

Business users often use data mining in a positive way in the form of Queries or Business Intelligence (BI) tools to provide a clearer line of site into their enterprise databases. It assists in the ability to drill down to discover losses, uncover anomalies, and to determine which areas are the more profitable parts of the business.

With respect to social networking, data mining can be used to harm an organization from the outside if employees are not made fully aware of the dangers of the information that is posted in status updates, blogs or profiles.

For example, LinkedIn is a popular social networking cloud application which is used as a professional networking site by its users for the purposes of maintaining virtual professional relationships, group membership, information sharing and job seeking. The profile format of LinkedIn allows users to share current and past work experience in a resume/CV format while linking these experiences to those of other LinkedIn users with similar experience in the same industries and companies. One of benefits of LinkedIn is that it helps its users take unstructured resume/CV data and enter it into a structured form for data validation. While the structured data benefit is helpful to users who wish to quickly find former colleagues, it also facilitates data mining for outsiders since users may not have considered their Linked In profile data can be used maliciously. The miner may gain a larger holistic informational picture of company data or critical information.

Data mining popular social networking sites by semi-skilled malicious outsiders could piece together data elements which may give them an insider view on information such as:

- The corporate technology infrastructure
- Data on lay-offs or downsizing
- Whether M&A activity is likely
- Pain points that a company might be experiencing including; financial stress, supply chain problems or specific customer service issues
- General sources of a company's human capital.
- Which competitors typically hire away a company's human capital.
- A list of current and past employees

DEFENSIVE STRATEGIES

Risky posts on social networks are an ongoing threat as these application sites are open and available 24 hours a day and users tend to create posts to them all of the time from a variety of devices and locations.

Defensive solutions and strategies need to be put into place and audited in order to protect the corporate enterprise against the risks associated with social networks. However, caution should be taken not to infringe on personal liberty should there be a need to use these systems. Organizations should seek the delicate balance of access and security by providing a foundational understanding of the risks to their users combined with proactive defensive strategies such as:

- Awareness
- Monitoring and Logging
- Acceptable Use Policy (AUP)
- Enforcement

AWARENESS

Making users aware of the risks associated with activity posting on social networks is an essential first step because awareness training is one of the best defenses against this activity.

Some of the risks associated with the use of social networking sites include: innocently disclosing

possible password clues, clicking on links to enable bots or malware, or disclosing information in status updates or profiles that can be data mined.

Recently, a questionnaire titled "How well do you know me?" was circulating on Facebook which contained questions like "What was your first pet's name?" or "What was the make and model of your first car?" Answering questions like these posted to your profile can later put you at risk for password hacking. These questions are typically password reset inquiries, or security questions for access to financial websites, as well as work log-in credentials, and touch upon other potentially damaging areas.

If an employee's individual workstation or log-in information becomes compromised by contracting a botnet [2] or backdoor bot, [13] then the risk is that such malicious payloads can spread to the entire network enterprise. This can potentially result in the theft of crucial and confidential company data or the destruction and the disruption of the enterprise infrastructure.

Social networks are proving to be a great venue for businesses to reach out to a wider customer and prospect base. [20] As businesses gather connections, friends and followers to their social network pages the risk increases.

These emerging social network risks should be incorporated into the Information Security Awareness training provided to company personnel.

ACCEPTABLE USE POLICY (AUP)

Once users are aware of the risks then, "an acceptable use policy" (AUP) [18] should be created and put into place. The AUP will provide the operational guidelines for posting company related data on social networks.

With respect to master accounts on social networks the AUP should identify who will be permitted to have access to the master account and what is acceptable to post. The AUP should also define procedures for reviewing and releasing data, governance of groups, job postings and protection of an organization's trademarks and logos.

Some companies may have existing policies that are similar in their "Media Policy and Procedure" documents. Infrastructure defenses must be reviewed periodically and should support the usage that is defined in the AUP. Procedures for logging in with authorization, and checks on those logs are necessary company defenses. Also, consider sharing the approved AUP policies with vendors who work on the company's behalf.

To begin, the meaning of "social networking" needs to be defined. Through this the users, management and vendor/partners understand what is considered a "social" site. The next step is to define "who" has access and/or authority to post on the organization's behalf. Most personnel should not be granted access to these sites since these sites are primarily for entertainment purposes. Allowing access may impede user productivity. Consideration may be given to marketing or HR, as these areas need to promote the brand, or seek out certain talent. Specialized service organizations may be employed to assist with the web presence. The concept behind an organization should be included, but avoid defining specific organizations as these may change.

Once the "who" is defined, the "what" should be reviewed, and a detailed analysis should be made. Policies should already exist concerning private business documents and these should be cross referenced or summarized within the AUP.

The more difficult task is assessing what information could be of use in data mining. This requires much forethought and "what if" scenarios. A few common areas to be cautious about are the organization chart, the listing of employees, and the job postings.

Organization charts, along with common email patterns (e.g. John.doe@myorg.com) may expose executives to un-solicited emails. Employee lists including job titles and the supervisory structure are perfect for staffing companies looking for top talent. Job postings for technical personnel can identify key components of the organization's infrastructure, useful for breaching a system.

Examining potential targeted stakeholders within the organization should be on the list, but the key is to be thorough. The AUP should list data-mining scenarios because this can give employees better insight and

may provide pause before they share crucial details. Having educated users will make the policy stronger.

In addition to routine information that is shared during work hours, the AUP must clearly define the information that is inappropriate to share at any time. If the organization is publicly traded, there may also be SEC regulations [6] to follow. Key personnel must receive strict instructions regarding the release of sensitive information. With a potential penalty of termination, persons associated with the organization must act appropriately and professionally.

MONITORING & LOGGING

Organizations should proactively monitor social networks for undesirable posting activity that could be data mined resulting in company harm. Hits involving unapproved social network activity should be recorded into an activity log for evaluation.

Monitoring and logging [7] activity on social networks is a challenge since activity can occur on many different and dynamically changing sites. Data is stored in many places and in varying formats.

The challenge is typically extended to consideration of internal and external oversight as corporate servers typically can only capture internal activity. Therefore, in order to detect and capture external activity from user owned devices and systems companies must deploy third party automated monitoring tools.

Automated monitoring tools are commercially available to provide proactive searches for predefined keywords in social networks. Examples of such tools include offerings by Visible Technologies and Radian6 which can be employed on a monthly or annual subscription basis. These sophisticated social media products, allow an organization to automate searches, monitor high volume keyword entries across many different channels, log events, run reports and provide automated alerts.

Activity logs are useful within an organization for documentation and purposes of future awareness training, and in some cases they may be needed for evidentiary purposes if the situation calls for it.

Internal social media monitoring and logging strategies should be designed and implemented by any organization whose employees are actively using social networks like Facebook, Twitter or LinkedIn from company managed equipment.

Furthermore, external social media monitoring and logging strategies should be designed and implemented by any organization whose employees are actively using social networks like Facebook, Twitter or LinkedIn from personally owned equipment.

Even companies which adopt a strict "no social network use policy" should perform external monitoring and logging of social networks to determine the chatter level about their brand within cyberspace.

ENFORCEMENT

At times certain activities may make enforcement of the AUP necessary.

Having an AUP in place without the complete support of upper management as well as proactive monitoring and logging is virtually impossible to enforce. [19]

The AUP should define acceptable use, violations or the "do not's" as well as possible corrective or disciplinary action which may result if the policy is violated. The AUP should also establish escalation of severity levels based on the type or category of unauthorized postings.

AUP enforcement may take on several forms including:

- Notification to an internal poster/employee that an unauthorized posting has been discovered on a social network site with a request to remove the offending post(s) within a prescribed number of hours/days.
- Notification to an external poster that an unauthorized posting has been discovered on a social network site with a request to remove the offending post(s) in a prescribed number of hours/days.
- An internal poster/employee may be referred to Human Resources if he/she continues to violate the AUP or generate an AUP violation that was categorized as a high a

severity level, thereby signifying a significant organizational risk.

- An escalation to legal council may be made if an internal or external poster continues to violate the AUP or if he or she were to violate the AUP with a high severity level.

To be efficacious an organization must maintain a proper chain of evidence of AUP violations for improper social network activity postings. Most of this evidence will be in the form of log files or automated alerts.

(Note the formatting from center to left on the following paragraph)

It is important to remember that a primary goal of the company's AUP is to deter unauthorized data from being disclosed on social networking sites and not to be unduly punitive with terminations or litigation.

CONCLUSION

Social networking sites are particularly useful in conveying messaging about a company and its brand, but a key goal must be to ensure the integrity of the organization including its vital infrastructure. In addition, improper information disclosures must be immediately removed or sanitized so that any negative risk is appropriately mitigated.

Clearly defined policies, tools and procedures to swiftly identify and remedy issues are crucial to provide a solid base for the desired protection. As with business continuity plans, it is of vital importance to monitor, test and adjust these procedures and tools, as necessary. Being vigilant with exploration of vulnerabilities and adjusting to meet these challenges will help to mitigate the risks that social networks pose to an organization.

Since harmful information can be produced outside the organization, engaging searches and monitoring software tools are necessary. Companies such as Facebook, LinkedIn and Twitter provide search utilities for "Social Media Monitoring" or "listening." These tools enable searches so that organizations can ensure proper messages are being passed, and allow them to respond to negative commentary. Along with these search engines, "Social Media Monitoring Tools" can provide automated alerts, and some can be integrated with CRM packages. Integration allows alerts to be routed to Sales, Customer Service or Marketing for appropriate research or response.

Active monitoring of log files is helpful to detect data leaks, as is having automated alerts warnings.

While enforcement is crucial, prevention is paramount to overall enterprise protection from an outsider's data mining.

REFERENCES

1. AMA (2010). AMA - Patient Confidentiality. Retrieved July 14, 2010, from American Medical Association: http://www.ama-assn.org/ama/pub/physician-resources/legal-topics/patient-physician-relationship-topics/patient-confidentiality.shtml
2. Barford, P.& Yegneswaran, V. (2006) An Inside Look at Botnets. Special Workshop on Malware Detection Advances in Information Security, Springer Verlag.
3. Barnes, S. (2006). A privacy paradox: Social networking in the United States. First Monday, volume 11, number 9.
4. Borders, K. & Prakash, A. (2009). Quantifying Information Leaks in Outbound Web Traffic. Proceedings of the IEEE 2009 Symposium on Security and Privacy, May 17-20, 2009, Oakland, CA, USA
5. Dillon, T. & Thomas, D. (2006). Knowledge of Privacy, Personal Use, and Administrative Oversight of Office Computers and E-mail in the Workplace. Information Technology Learning and Performance Journal, volume 24; number 2, pages 23-34.
6. Elliott, J., Morse, D. & Richardson, G. (1984) The Association between Insider Trading and Information Announcements. The RAND Journal of Economics, volume 15, No. 4 (Winter, 1984), pp. 521-536
7. Fawcett, T. & Provost, F. (199) Activity monitoring: noticing interesting changes in behavior, Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, p.53-62, August 15-18, 1999, San Diego, California, United States
8. Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996) From data mining to knowledge discovery: an overview, Advances in knowledge discovery and data mining, American Association for Artificial Intelligence, Menlo Park, CA
9. Gross, R., & Acquisti, A. (2005). Information revelation and privacy in online social networks. Proceedings of the 2005 ACM workshop on

Privacy in the electronic society. Alexandria, VA USA, Pages: 71 – 80, ISBN: 1-59593-228-3.

10. Hricik, D., (2010) Communications and the Internet: Facebook, E-Mail, and Beyond. Available at SSRN: http://ssrn.com/abstract=1557033
11. Kolek, E., & Saunders, D. (2008). Online Disclosure: An Empirical Examination of Undergraduate Facebook Profiles. Journal of Student Affairs Research and Practice, volume 45, issue 1.
12. Kursawe , K. & Katzenbeisser, S. (2007). Computing under occupation, Proceedings of the 2007 Workshop on New Security Paradigms, September 18-21, 2007, New Hampshire.
13. Naraine, R. (February 2007) Is the botnet battle already lost? Available at http://www.eweek.com/print_article2/0,1217,a=191391,00.asp.
14. Narvaez, J., Endicott-Popovsky, B., Seifert, C., Aval, C. & Frincke, D. (2010) "Drive-by-Downloads," hicss, pp.1-10, 2010 43rd Hawaii International Conference on System Sciences.
15. Northcott, S. (2010). Apple Hiring Antenna Engineers Replacement iPhone 4 fixes issue for one. Touch Reviews. Available at http://touchreviews.net/apple-hiring-antenna-engineers-replacement-iphone-4-fixes-issue
16. Orgill, G., Romney. G., Bailey. M. & Orgill, P. (2004). The urgency for effective user privacy-education to counter social engineering attacks on secure computer systems, Proceedings of the 5th conference on Information technology education, October 28-30, 2004, Salt Lake City, UT, USA
17. Siponen, M. (2000). A conceptual foundation for organizational information security awareness, Information Management & Computer Security, volume 8, Number: 1, pp.31 – 41
18. Stewart, F. (2000). Internet acceptable use policies: Navigating the management, legal, and technical issues. Security Management, (July/August), 46-52.
19. Volonino, L. & Robinson, S. (2003) Principles and Practice of Information Security, Pearson Education.
20. Zagorski, D. (2009). The Three Most Popular Social Networks for Business (and Why You Should Use Them). LeftBrainRightBrain Marketing. Available at http://lbrbmarketing.com/greymatter/pdf/1109-which-social-network.pdf