2007

BLOGS: ANTI-FORENSICS and COUNTER ANTI-FORENSICS

Glenn S. Dardick
Longwood University

Claire R. La Roche
Longwood University

Mary A. Flanigan
Longwood University

Originally published in the Proceedings of the 5th Australian Digital Forensics Conference, Edith Cowan University, Perth Western Australia, December 3rd 2007.

BLOGS: ANTI-FORENSICS and COUNTER ANTI-FORENSICS

Glenn S. Dardick
dardickgs@longwood.edu
Longwood University

Claire R. La Roche
larochecr@longwood.edu
Longwood University

Mary A. Flanigan
flaniganma@longwood.edu
Longwood University

Abstract

Blogging gives an ordinary person the ability to have a conversation with a wide audience and has become one of the fastest growing uses of the Web. However, dozens of employee-bloggers have been terminated for exercising what they consider to be their First Amendment right to free speech and would-be consumer advocates face potential liability for voicing their opinions. To avoid identification and prevent retribution, bloggers have sought to maintain anonymity by taking advantage of various tools and procedures - anti-forensics. Unfortunately some anonymous bloggers also post content that is in violation of one or more laws. Some blogging content might be viewed as harassing others - an area known as cyber-bullying. Law enforcement and network forensics specialists are developing procedures called Counter Anti-forensics that show some promise to identify those who violate the law. However, these techniques must be used with caution so as not to violate the rights of others.

Keywords

digital forensics, anti-forensics, counter anti-forensics, blogs, stylometrics

INTRODUCTION

In 2006, Time magazine's Person of the Year was "You". Blogging was in part, responsible for that choice. Time's explanation was that the Web was being revolutionized and used as "a tool for bringing together the small contributions of millions of people and making them matter" (Grossman 2006). Recent surveys conducted by the Pew Internet & American Life Project indicated that approximately 39% of adults in the U.S. read blogs and 8% of Americans participate in this form of personal publishing (Lenhart and Fox 2006).

Time saw the Web as "an opportunity to build a new kind of international understanding, not politician to politician, great man to great man, but citizen to citizen, person to person. It's a chance for people to look at a computer screen and really, genuinely wonder who's looking back at them." (Grossman 2006) As it turns out, many of those who blogged now know who has been looking back at them and perhaps wished their contributions could have been made anonymously, and kept so.

RETRIBUTION

While many blogs are frequently posted without consideration of the appropriateness of the contents, others are posted fully cognizant that the content may be inappropriate or offensive. (Barnes 2007, Howell 2007, Williard 2007) There appears to be a tendency to reveal information or express thoughts in a blog that one would be reluctant to say in person or in a traditional print medium. This is particularly true if the blogger believes that s/he is blogging anonymously.

According to a 2007 Proofpoint survey, approximately 1 in 10 companies have fired an employee for blogging or message board postings (Proofpoint). Matthew Brown, a former Starbucks' employee in Toronto, was fired for mentioning in one posting that his supervisor did not let him go home when he was sick. (Koulouras 2004). Another blogger, Heather Armstrong, the original "dooced" (terminated for blogging) employee found out regarding blogs and employers that "They specifically will find it and read it, and all hell will break loose." (Witt 2004) Armstrong was fired from dooce.com in 2002 when her employer found the contents of her blog to be offensive.

Blogging is by definition a public activity and as such there should be no reasonable expectation of privacy. In fact, many tools such as Google's Blog Search and Really Simple Syndication (RSS) feeds make it very easy to

find and read specific blog content. Although bloggers have a First Amendment right to express their opinions, they are not protected from the consequences of such expressions.

ANONYMITY AND ANTI-FORENSICS

Blogging and Anonymity

To avoid identification and possible retribution, some bloggers will attempt to remain anonymous. However, to maintain anonymity on the Internet, bloggers must hide not only "who" they are, but "where" they are, and "what" equipment they are using. This is all information that may readily be obtained by accessing the blog site.

A blogger's identity can be determined easily through payment and/or registration records. A blogger's identity is at risk of being exposed when the blogger uses their credit card or real name in paying for, or registering a website or e-mail address. Bloggers can prevent this by obtaining an e-mail address that provides anonymity and paying with a Virtual debit/credit card that cannot be traced back to the purchaser.

A blogger might be located through the IP address that was assigned to the computer used at the time the blogger accesses the site via the Internet. To keep their IP address anonymous a blogger can go through an intermediary on the Internet referred to as a proxy. Proxies may be used to access other proxies referenced by proxy systems such as Tor (Tor 2007). The Tor proxy system randomly selects a chain of proxies from an inventory of available proxies provided voluntarily by users of the Tor system. The system supplies proxies from multiple countries.

Information specific to the blogger, such as the operating system and browser being used, is passed through the Internet via "Headers" when access to the blog is made via the Internet. Additional information may be gathered from the blogger's system if scripts are allowed to run within the blogger's browser. To avoid passing information that might identify details of the blogger's system, "Headers" from the blogger's system, can be dynamically altered via proxies, or transcoders, capable of filtering and modifying such information. Potential threats from scripts can also be blocked via the use of filters from such proxies. Filters are capable of changing or completely removing code which can compromise a system's and/or blogger's identity. Filtering capability is provided by proxy products such as Privoxy (Privoxy 2007) that are readily available on the Internet (Privoxy 2007, Tor Download 2007, Torbutton 2007, and Vidalia 2007). Transcoding has previously been used to selectively optimize bandwidth by converting graphics files (Han 1998). Transcoding can also be used to replace HTML statements to allow web content to be displayed more appropriately on a wider variety of devices including hand-held devices. While transcoding can be server-based it can also be client-based or proxy-based using tools such as Proxomitron (Lemmon 2007). Originally, Scott Lemmon's Proxomitron was meant as a way to dynamically modify HTML to remove advertisements and pop-ups. Proxomitron eventually grew to become a general purpose proxy-based transcoder that could be a client-based or a server-based proxy. Much of the software is readily available as downloads from the Internet along with ample advice, recommendations and support from organizations and individual websites (Electronic Frontier Foundation 2005, Morris 2005, Zuckerman 2007).

Harmful Blogging: Juror Misconduct, Cyberbullying and Cyberstalking

Unfortunately, sometimes blogging can run counter to the law and/or cause harm. In some cases, the blogger may be ignorant of the illegality of their actions and may or may not try to conceal their identity. Blogging has also become an avenue for both Cyberbullying (Williard 2007) and Cyberstalking (Barnes). Several cases have occurred recently where blogging has been used to inflict harm, some resulting in death (Taylor 2007). In other cases, jurors have ignored instructions from the court and have not only discussed the case prior to deliberations, but have actually written about the cases in blogs (Howell 2007). Perhaps even more insidious is when someone, such as a witness, tries to influence a jury and public opinion through information published on the Internet. In such cases, the witness would attempt to remain anonymous. Only if it could be proven that it was, in fact, a witness or other person directly involved in the proceedings, would there possibly be consequences resulting in a mistrial. It is possible that further evidence may be uncovered indicating the motives as to why a witness would be trying to influence a jury above and beyond what their testimony would provide.

Recently, President Bush signed the "Violence against Women and Department of Justice Reauthorization Act of 2005" (H.R. 3402). In effect, the bill modified 47 U.S.C. §223 to read "Whoever...utilizes any device or software that can be used to originate telecommunications or other types of communications that are transmitted, in whole or in part, by the Internet... without disclosing his identity and with intent to annoy, abuse, threaten, or harass any person...who receives the communications...shall be fined under Title 18 or imprisoned not more than two years, or both." (McCullagh 2006)

COUNTER ANTI-FORENSICS

Counter Anti-Forensics may be used to determine the identities of bloggers who attempt to remain anonymous. Such techniques may be as basic as determining if the blogger has hidden all of the identifying tracks. For instance, a blogger may have acquired a false e-mail address and applied a fictitious name; however they might not have hidden their IP address (Zuckerman 2005). If a blogger successfully takes all of the necessary steps to hide their whereabouts, their name, and all identifying information about their equipment, what is left to identify the blogger is the content. There are methods that may tie the content of the blog to an individual, but the reliability may not be sufficient in a court of law. It may, however, be sufficient to establish probable cause and to acquire a warrant to search for additional information on the suspect's equipment. Such methods utilize stylometrics for determining author attribution.

Stylometrics

Stylometrics is defined as "techniques used for the quantitative examination of textual styles, often used as a method for authorship attribution studies." (AHDS 2007). The research into stylometrics has resulted in its application within forensics examinations. One such method of author attribution, QSUM or CUSUM, was developed by Jill M. Farringdon (Farringdon 1996). It creates, in effect, a cyber "fingerprint" for an author. Cyber "fingerprints" have been researched and referenced by Li and Chen (2006) as "Writeprints". The work is based in part on earlier research into author attribution of e-mails (de Vel, Anderson, Corney and Mohay 2001). Cyber fingerprints can be classified into four categories: lexical, syntactic, structural, and content-specific (Abbasi and Chen 2005).

Lexical attributes include characteristics such as total number of words, words per sentence, word length distribution, vocabulary, total number of characters, characters per sentence, characters per word, and the usage frequency of individual letters. Syntax attributes refer to the patterns used to form sentences such as punctuation. Structural attributes refer to the text's organization and layout as well as font, hyperlink, and embedded image characteristics. Content-specific attributes are words that may be important or have special relevance within a domain (Abbasi and Chen 2005). Unfortunately, the ability to disguise authorship of electronic communications through imitation, and techniques such as cut and paste, is potentially high (De Vel 2001). However, stylometric evidence has been admitted in court, passing both the Daubert and Frye criteria (Chaski 2005).

In one recent case, *Connecticut v. Julie Amero*, a detective who was the investigator in the case was active in communicating his "case" to the public (Bass 2007). There were similarities in those communications and several blogs that were posted anonymously while the defendant was awaiting sentencing of up to 40 years. Because of the sensitivity of the case at that moment, the blogs were not further analysed, and the verdict was in fact thrown out for unrelated reasons. Had the verdict not been thrown out at sentencing and a tie-in between the blogs and the detective established, the legal ramifications might have been very enlightening,

CONCLUSION

As the use of anti-forensics methods increases, the application of counter anti-forensics methods will increase as well, going beyond traditional Digital Forensics methods and incorporating stylometrics-based methods to assist in determining authorship. In fact, such methods have been deemed sufficient in a court of law to determine authorship. Results of digital forensics methods might also result in the ability to acquire warrants and enable law enforcement personnel to retrieve additional evidence in an investigation.

A significant amount of research has been done in the areas of stylometrics and author attribution. Much of the research is applied in determining whether certain material is likely, or not, to be attributable to a specific author. While much of the early research focused on literary works, these methods are now being applied to blogging posts and other electronic communications such as e-mail. Much of the research uses a defined set of authors to determine attribution. This research results in attempting to show attribution of a specific document(s) to a specific author from a defined set. More research is necessary to determine how stylometrics should be used within the digital forensics process models and the level of certainty required to show probable cause, reasonable suspicion and/or obtain warrants.

There is a need for a closer tie between stylometrics and investigations employing digital forensics. Many of the digital forensics process models start with a specific suspect and evidence potentially related to that suspect. Thus, the role has been one of confirmation rather than identification of a suspect. The process models need to look at piercing the shield of anonymity in order to reasonably identify potential suspects and discover evidence. The question for the future is what role can the digital-forensics process play in light of anonymity? Can it be effective, not simply after the filing of charges or the issuing of a warrant, but prior to the identification of a specific suspect?

REFERENCES

Abbasi, A. and Chen, H. (2005). Applying Authorship Analysis to Extremist-Group Web Forum Messages. IEEE Intelligent Systems 20(5):67-75

AHDS (2007). “Stylometrics”. http://ahds.ac.uk/ictguides/methods/method.jsp;jsessionid=4F089923B357BA5BFA77FEF1C1374391?methodId=64

Barnes, S and Biros, D. (2007) “An Exploratory Analysis of Computer Mediated Communications on Cyberstalking Severity”. Journal of Digital Forensics, Security and Law. 2(3):7-27

Bass, S. (2007). “Detective Speaks Out in Teacher Porn Case”. http://blogs.pcworld.com/tipsandtweaks/archives/003745.html

Chaski, Carole E. (2005), "Who’s At The Keyboard? Authorship Attribution in Digital Evidence Investigations". International Journal of Digital Evidence, Spring 2005, 4(1)

de Vel, O., Anderson, A., Corney, M., and Mohay, G. 2001. Mining e-mail content for author identification forensics. SIGMOD Rec. 30, 4 (Dec. 2001), 55-64. DOI= http://doi.acm.org/10.1145/604264.604272

Electronic Frontier Foundation (2005) http://tor.eff.org/eff/tor-legal-faq.html.en (accessed September 5, 2007).

Farringdon, J. (1996). “QSUM The Cumulative Sum (cusum) Technique for Authorship Analysis & Attribution”. http://members.aol.com/qsums/

Firefox http://www.mozilla.com/en-US/firefox/ (accessed September 5, 2007).

Grossman, Lev (2006), “Time’s Person of the Year: You”, TIME http://www.time.com/time/magazine/article/0,9171,1569514,00.html

Han, R. et al., Dynamic Adaptation In an Image Transcoding Proxy For Mobile Web Browsing in IEEE Personal Communications, Dec 1998, 8-17.

Howell, D. (2007). “Blogging jury duty”. http://blogs.zdnet.com/Howell/?p=104 (accessed November 16, 2007)

H.R. 3402. President Signs H.R. 3402, the "Violence Against Women and Department of Justice Reauthorization Act of 2005". http://www.whitehouse.gov/news/releases/2006/01/print/20060105-3.html

Koulouras, Jason (2004) “Employee fired by Starbucks over Blog”, Blogcritics magazine, http://blogcritics.org/archives/2004/09/04/141004.php (accessed July 29, 2007).

Lemmon, S. (2007) “Proxomitron”. http://www.proxomitron.info/ (accessed November 16, 2007)

Lenhart, Amanda, Fox, Susannah (2006) “Bloggers A portrait of the Internet’s new storytellers”, Pew Internet & American Life Project, July 19, 2006.

Li, J., Zheng, R., and Chen, H. 2006. From fingerprint to writeprint. Commun. ACM 49, 4 (Apr. 2006), 76-82. DOI= http://doi.acm.org/10.1145/1121949.1121951

McCullagh, D. (2006). “FAQ: The new 'annoy' law explained” . http://www.news.com/FAQ-The-new-annoy-law-explained/2100-1028_3-6025396.html (accessed November 16, 2007)

Morris, Sofia (2005), “An Anonymous Blogger Tells All”, http://journalism.nyu.edu/pubzone/notablog/story/anonymous/ (accessed September 5, 2007).

Privoxy http://www.privoxy.org/ (accessed September 5, 2007).

Proofpoint, Inc. (2007) “Outbound email and Content Security in Today’s Enterprise” www.proofpoint.com

Taylor, B. (2007). “Mom: Web Hoax Led Girl to Kill Herself” http://ap.google.com/article/ALeqM5gg5xCtQtLBF6vJqWXStItGEOsJfwD8SV6U680 (accessed 11/16/2007).

Tor http://tor.eff.org/overview.html.en (accessed September 5, 2007).

Tor Download http://tor.eff.org/download.html.en (accessed September 5, 2007).

Torbutton http://freehaven.net/~squires/torbutton/ (accessed September 5, 2007).

Vidalia http://vidalia-project.net/index.php (accessed September 5, 2007).

Williard, N. (2007) “Educator’s Guide to Cyberbullying and Cyberthreats”, http://cyberbully.org/cyberbully/docs/cbcteducator.pdf (accessed November 16, 2007)

Witt, April (2004), “Blog Interrupted”, Washington Post, August 15, 2004, p. W12.

Zuckerman, Ethan (2005) Global Voices http://www.globalvoicesonline.org/?p=125 (accessed September 5, 2007).

COPYRIGHT