2015

Mining social networking sites for digital evidence

Brian Cusack
Auckland University of Technology

Saud Alshaifi
Auckland University of Technology

DOI: 10.4225/75/57b3f23afb885
13th Australian Digital Forensics Conference, held from the 30 November – 2 December, 2015 (pp. 15-21), Edith Cowan University Joondalup Campus, Perth, Western Australia.

MINING SOCIAL NETWORKING SITES FOR DIGITAL EVIDENCE

Brian Cusack; Saud Alshaifi
Auckland University of Technology, Auckland, New Zealand
{brian.cusack; saud.alshaifi} aut.ac.nz

Abstract

OnLine Social Networking sites (SNS) hold a vast amount of information that individuals and organisations post about themselves. Investigations include SNS as sources of evidence and the challenge is to have effective tools to extract the evidence. In this exploratory research we apply the latest version of a proprietary tool to identify potential evidence from five SNS using three different browsers. We found that each web browser influenced the scope of the evidence extracted. In previous research we have shown that different open source and proprietary tools influence the scope of evidence obtained. In this research we asked, What variation in the scope of evidence extraction can be expected between different browsers? The implications of this exploratory research is for precaution. The choice of a web browser used to investigate a SNS directly influences the scope of digital evidence obtained.

Keywords
Social, Networking, Investigation, Browsers, Evidence

INTRODUCTION

Online social networks are forums hosted on the Internet that provide easiness and effectiveness for unlimited amounts of users to share information in digital forms such as images, texts, links, audios, and videos. In simple terms they are massive communication platforms (Cheung and Lee, 2010). The use social media has become pervasive in the lives of many users and an extension of their real lives. There are many different online social networks with different purposes of use but all communicate information about the individual, the organisation and their networks of association. The largest one Facebook had over 1.5 billion users in May 2015 (Chen, Xu, Yuan, and Shashidar, 2015). Many users of these sites have become psychologically attached to the interaction and the self-promotion to a point where they freely post information about themselves, including pictures, status, comments, locations, beliefs, opinions and feelings. Some of these communications may be exaggerations or fabricated using information tools but many users are simply conveying stories in various forms about themselves, the organisation and their communities ( ). The nature of the medium provides a sense of security and personal safety in which the constraints of normal social settings are often absent. In this context much information is available that would not always be accessible by standard investigation techniques such as interviews and observations. In a standard interview situation cues are present that can inhibit or facilitate conversation and the recollect of events. The proximity of relationships in such situations often colours the tone and texture of what is recorded. In the online situation the spatial proximity of others may be distant or non-existent. The user conveys their own ideas and beliefs about matters in a mobile non-space with only the influence of the online community and perceptions of others built up from previous experience of online interaction. In some instances the user is conveying information on behalf of others such as business advertising or communications for those without an account but in every instance the user mediates the situation. As such the user is in a position of power, can earn trust and can act as an authority in matters unknown. In simple terms the user feels important and capable to declare truths (Wang, Woo, Lui, Quak, Yang, 2012).

From a forensic point of view, online social networks are a potential source of evidence that can help during investigation (Yue et al., 2014). There are many instances where employers, statuary authorities and others have consulted social networking evidence before making judgements (Zainudin et al., 2010). Organisations have also encouraged members to maintain social network sites to communicate and promote brand (Dar and Shah, 2013). Many work performance appraisals include both positive and negative feedback based on online social performance (Mingming, 2014). These feedback are based on “likes”, numbers of followers, numbers of retweets, and so on. As a consequence there is a huge amount of data available online surrounding individuals and any other created entity such as political parties, brands, clubs, charities, and so on. The access to some of the data is controlled by privacy settings and privileges but in general there is an enormous amount of data available. Not surprisingly searching and sorting through the entangled medium is a big part of digital investigation. Social engineering has also played a role in weakening the protections of privacy and often social networks that are perceived to be secure are only as strong as the weakest link who may decide to disclose sensitive information. Publically reported cases of mass hacking of passwords has led to the disclosure of sensitive documents, images and audio clips that were only intended for the

select social group (Zhang, Choudhary, and Grudin, 2014). All of these matters show that online social networks contain a multiplicity of evidence that can be accessed in many ways. The evidential value is moderated by the potential for over-inflated claims, digital make-overs and flaring; but the amount of data available and the improvement in extraction and analysis tools is making the source unavoidable.

This paper is structured to review previous literature, define the test set up and to report the results. The focus of the research is to identify the effect the use of a particular web browser has on the scope of evidence. The results are then discussed to elicit the implications for practice. The conclusion is that social networking site (SNS) digital forensic tools have to be used with caution and the interpretation of findings moderated to allow for variations caused by web browser effects and tool effects.

PREVIOUS LITERATURE

The introductory section has defined and described the characteristics of SNS. This section is concerned with the identification of evidence that may be found on a SNS. The opportunity to collect evidence from SNS is no different than any other digital forensic collection activity and it must comply with standardised criteria for acceptance. The five Daubert criteria are often cited as helpful guidelines for assuring evidence (Cohen, 2010). Although these are certainly helpful overriding guidelines for technical performance, legislation in the form of a jurisdiction Evidence Act, for example, may take priority. If the evidence has been collected in an unacceptable social sense where coercion has been used or reasonable privacy breeched; or where spoliation may have occurred, then the evidence becomes unacceptable (Jang and Kwok, 2014). The investigator has to comply both with the IT technical criteria, the legal framework and due processes to assure admissibility. In the following two sub sections the previous literature on SNS tool testing is reviewed and the type of digital evidence available summarised.

TOOL EVIDENCE COLLECTION

In previous literature variations between tool performances for extracting digital evidence from SNS have been reported. The findings show that the capability of three widely available tools for evidence extraction in SNS varied greatly under test conditions. The experiment was set up using standardised tool testing algorithms (see Table 1). The capabilities of digital forensic tools to perform examination and extraction of social networking artefacts was measured. The overall top performing tool was EnCase Forensic which was able to automatically examine and extract an average of 89.6% of all SNS artefacts. It also had a perfect rating of 3/3 for all test cases. The comprehensive Internet history search was a useful technique to isolate SNS related artefacts and in two scenarios the provided scripting language EnScript was use to code and facilitate the extraction.

Table 1. Test Cases for Forensic Tool Testing

Test Case #	Test Case Name	Tested Tool Functionality
TC01	SNS History Analysis	Provide detailed list of SNS URLs accessed.
TC02	Web Browser Cache Analysis	Automatically examine and decode Web browser cache for SNS information, data and files.
TC03	SNS Session Analysis	Locate Internet session artefacts created by SNS interaction.
TC04	Facebook Chat Analysis	Automatically examine evidence for Facebook chat messages.
TC05	Repeatability & Reproducibility	Tool achieves same results consistently.

CacheBack also performed well at automatically examining SNS artefacts, with an average of 75.2% of all artefacts extracted. Internet Evidence Finder (IEF) is designed to be run on a live system, and did not process static forensic image files as used in the testing. The result was a poor performance and the requirement to mount the forensic image for evidence extraction. These results all have implication for evidence scope. Each of the three tools also faced challenges in the area of web browser compatibility and interoperability. Table 2 summarises the performance findings.

Table 2. Summary of Findings: Automated Examination Capabilities of Tools

Scenario	CacheBack		IEF		EnCase	
	Score	Rating	Score	Rating	Score	Rating
TC01	**58.2%**	**2**	**21.4%**	**1**	**93.2%**	**3**

TC02	**31.8%**	**1**	**0.6%**	**1**	**87.8%**	**3**
TC03	**86.1%**	**3**	**0%**	**0**	**92%**	**3**
TC04	**100%**	**3**	**50%**	**2**	**75%**	**3**
TC05	**100%**	**3**	**100%**	**3**	**100%**	**3**
Total Weighted	**75.2%**	**12**	**34.4%**	**7**	**89.6%**	**15**
Ranking	**2nd**		**3rd**		**1st**	

The Type of Evidence Stored In A SNS

The type of data that can be found in SNS is impressive. We have listed categories in previous articles but a short review would include the following. Five working categories are (Mumba and Venter, 2014):

Service data: Data that has to be provided by users to continue using the social network site, examples of data is legal name, Date of Birth, and phone numbers and so on.
Disclosed data: Any data posted by the account user, it could be presented in any format such as pictures, videos, links, comments, and updating status.
Entrusted data: Any data posted by someone else to a user account (friends, subscribers, followers etc.). The difference between Disclosed and Entrusted data is that the user does not have control over the entrusted data once it's been posted.
Incidental data: Incidental data is what other people write in their account about a particular user. Again the data could be anything, pictures, messages, videos, and so on.
Behavioural data: The data collected by the site about users' practises and habits. By recording their activities, choices, regular habits, points of views and so on.

The types of evidence can vary from one SNS to another depending on their architecture and the features provided. The different data sources that that can be collected from online social networks, which may lead to acceptable evidence during a forensic investigation can be grouped into five areas (Zainudin, Merabti and Jones, 2010).

Social footprint with other users, including friend lists, connected groups, who are the followers, and following who.
Communication methods between the users within the site, e.g. private messages, instant messenger, comments, likes, group communications, and events.
Pictures and videos posted by the users, and who were tagged in the pictures, what other pictures a certain user was tagged on.
The times of activities: when a specific user logged on into the site, and what sort of activities were performed in a specific time frame.
The applications used by the user, and identifying the purpose of the used apps, and what information can be deduce in the social context.

The evidence may be spread in a variety of locations some of which may be inaccessible. Online social networking sites exploit the services of cloud providers leaving potential sources of information lost in the complexity of commercial arrangements and service level agreements. A traditional investigation would have relied on browser forensic techniques to access a large percentage of the potential evidence on a user's hard drive. However the newer information architectures suggest that RAM dumping and cloud investigation techniques are also required.

TEST SETUP

The research is to investigate the variation in scope of different web browsers to extract digital evidence. The test context was the University laboratory using one PC and the network that is available to all University users. Each web browser was checked for the standard out of the box configurations before use. In this way we sought to generalise the experience any user may have on the network when accessing the five SNS with the three different web browsers. The test environment was set up using accounts created for the purpose of the research on five different SNS (Facebook; Twitter; Instagram; Linkin; Bayt). Consequently the data reported is all fictitious scenario based and with no real owners. Three web browsers were tested to observe the scope of evidence extraction. The web browsers were selected from the four most popular web browsers in 2015 (Statcounter.com, 2015). The popularity rank was reported as Chrome (50%), Internet Explorer (IE) (12%), Firefox (10%), Safari (13%). Consequently Chrome, Firefox and IE were selected based on availability and platform. The techniques from

Browser Forensic investigation were employed, such as RAM dumping, Cache analysis from the PC, and so on. In figure 1 the phases of investigation are shown to communicate the scope of enquiry. The focus was on site evidence availability and web browser extraction capability.

The researchers played case scenarios to populate the sites with potential evidence and also to control the amount and type of data for discovery before the tests began. In this way the web browsers could be assessed against the potential evidence present and ranked. In the first phase the IT artefacts were defined and the data types identified and posted. The second phase determined the tests and scenarios. The evidence was subsequently stored on a hard drive and in the PC 8Gb RAM. In phase three the hard drive was imaged and the RAM dumped for analysis in Phase 4. Phase 5 was the research report. The Third phase acquisition and extraction of evidence was performed based on NIST forensic best practice guidelines. The Fourth phase included a reconstruction of the evidence against the research scenarios (see figure 1). The following hardware and software were used:

- Darik and Nuke wiping utility tool for zeroing the suspect hard drive before conducting the scenarios (Generating the controlled data).
- Windows 7 professional with 32-bit Operating System is installed on the suspect computer, and latest versions of IE, Chrome, and Firefox.
- FTK imager light 3.1.1: used for acquiring RAM memory of suspect's computer.
- AccessData® FTK® Imager 2.9.0.138: used for creating an image of memory dump and the acquired hard drives and to verify the integrity of the image by calculating MD5 and SHA1 values.
- Tableau eSATA forensic bride: A forensic SATA/IDE bridge model T35es is used to acquire computer hard drive in forensic manner where the evidence is not altered or changed.
- Tableau Imager 1.11: To acquire the computer's hard with the use of tableau eSATA forensic bridge.
- Belkasoft Evidence Center 7.3: used for analyzing OSNs evidence from the suspect's computer.

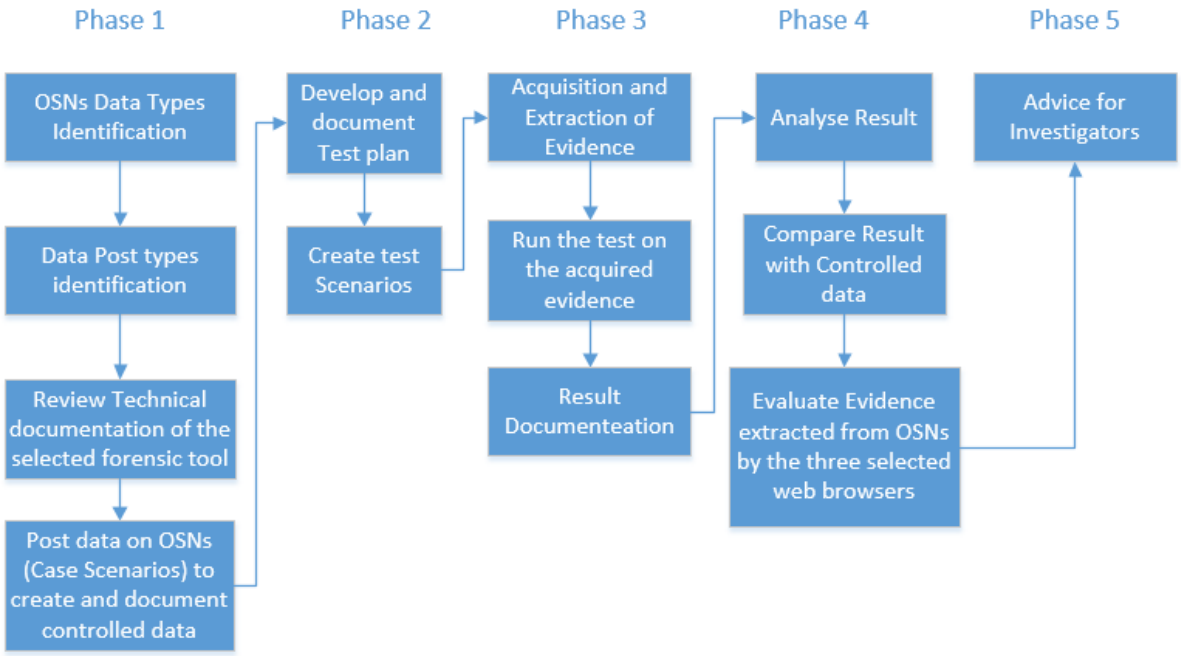

Figure 1. Research Phases

THE RESULTS

The experiment confirms that the types and amounts of data that can be found in SNSs is large. It is also comprehensive and descriptive. However, the results show that not all data can be accessed and that different web browsers find different and different amounts of data. In the experiment some of the posted data was not found at all. The RAM, pagefile.sys, and hard drive were all searched for evidence and because we had set the experiment up knowing what could be found the web browser limitations became explicit. The figure 2 summarizes the percentage

of forensic evidence found with each web browser. The data is further differentiated into SNS (colour code) and location.

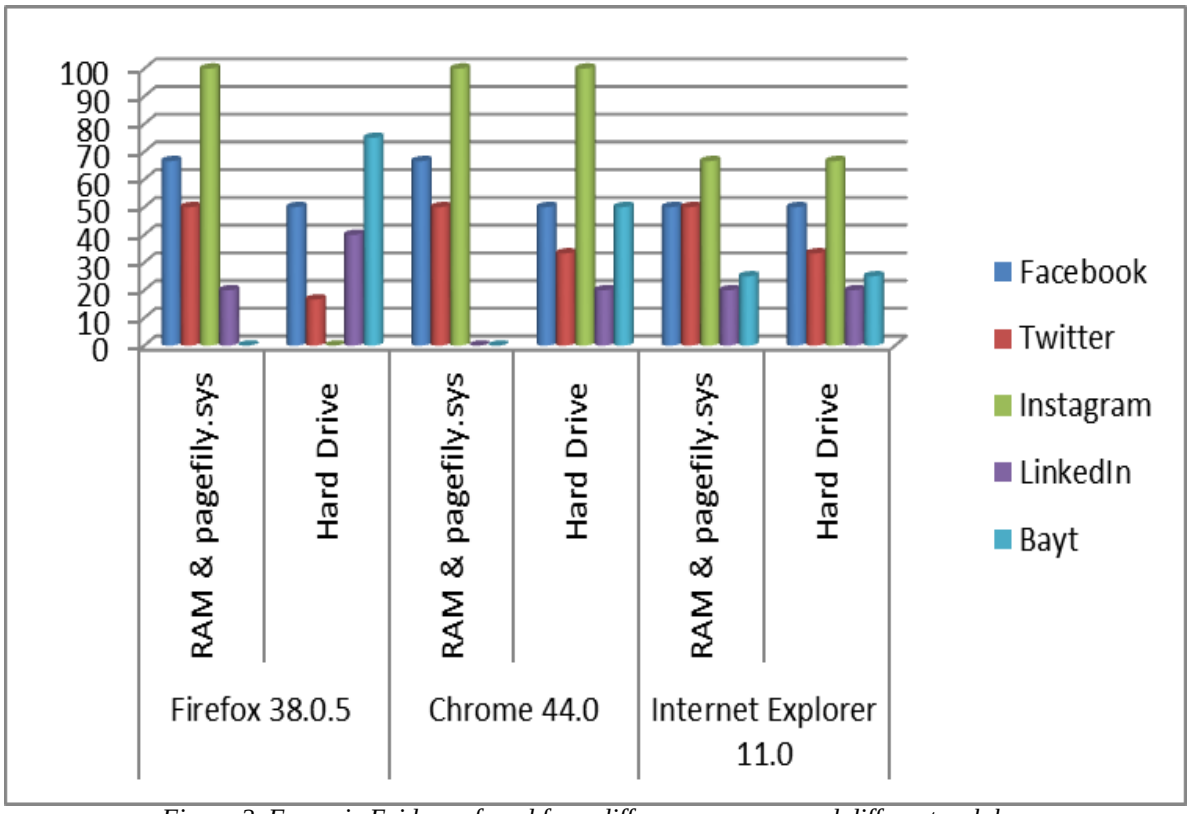

Figure 2. Forensic Evidence found from difference sources and different web browsers

DISCUSSION

The results in figure 2 show considerable variation between web browsers and SNS when evidence discovery is required. Each web browser was configured and used in the fashion described in the set up section. Across the three web browsers Firefox and Chrome performed effectively for the SNS Twitter but did not show any remarkable performances for the other four SNS. Internet Explorer similarly stood out as underperforming on all SNS. It can be argued that Internet Explorer was more consistent across the five SNS than the other two web browsers but in terms of a high 95% or better discovery rate, it fell well short. The results from the study have to be read with the limitations described in the set up section and the best advice to a Digital Forensic Investigator is cautioned with these limitations. In practice the results suggest that a range of web browsers should be used to provide the best scope of discovery and Internet Explorer can provide a baseline from which to work. The consistency across SNS and storage devices suggests that affair indication of where to look may be established. Once the baseline is set then a Digital Forensic Investigator can use other web browsers to search more thoroughly in the target areas.

The tools a digital investigator uses for extracting evidence from a SNS determine the scope of the evidence acquired. Previously published experimental results show that the choice of evidence extraction and analysis tools sets determine what is found from a SNS. In this research the experiment shows that the scope is further influenced by the type of SNS, the web browser and where the evidence is stored. The results showed that the easiest SNS to extract evidence was Instagram and the hardest LinkedIn. Other extreme variations were noted when Firefox and Chrome web browsers were used on the Bayt SNS. The extraction of evidence from different locations is consistent across the RAM, pagefily.sys and hard drive for Internet explorer but variable for Chrome and Firefox. This suggests that again the web browser used by the investigator is to impact the scope of evidence obtained. The clear message from figure 2 analysis is that with the exception of Instagram over 50% of the structured evidence was not extracted by any of the three web browsers in the experiment. This is a very large proportion and of material influence when reporting evidence obtained from a SNS. A professional investigator is to demonstrate in practice the scientific methods used to extract the evidence so another can extract the same evidence using the same methods. However it is problematic that if another expert using the same web browser with the same configurations would find the same evidence or if a different web browser performed the same method the same evidence would be extracted.

The implications of these findings is for further caution and preparation by investigators when collecting evidence from SNS. The technical report of findings has to include a specification of the web browser(s) used and the version numbers so that another expert can best chance replicate the method and procedures to obtain the same evidence. More worrying however, is the potential for Judges and Juries to simply take evidence extracted from SNS at face value and assume it has the same status as digital evidence from traditional sources. SNS introduce extra layers of complexity that may not yet be fully understood and planned for in digital evidence extraction. The extent and limitations currently found in digital evidence extraction tools and web browsers can seriously limited the scope of what may be obtained. An incomplete picture of an event may provide items of interest but cannot tell the whole story. In addition the way people can behave online places limitations on the value of the evidence. Previously mentioned behaviours such as flaring, bragging, serialising, fantasising and so on limit the reality factor associated with any posted data. Such semantic slippage and detached referrals may fuel a defence case that alleges planting of evidence and the fabricating of charges. The worst case situation leaves an SNS user vulnerable to a multiplicity of theories of the evidence and consequently accountable to unrelated accusations. The implications of these findings is for caution when extracting and reporting evidence from SNS. An investigator may use multiple tools and web browsers but still has to accept that a comprehensive search of a SNS for evidence remains incomplete.

CONCLUSION

There are many cases where people have used online social networks to reveal their admission of committing offenses. Often the motivation is to brag or to seek popularity. The social networks people disclose online also have the links to others who influence their behaviour and those with whom they exert influence. Of course there are many strengths of relationship within any online social network but the nature of trust and apparent removal of the usual barriers to expression allows the disclosure of important information. In this research we show that caution must be taken when evidence is extracted and reported from SNSs. The performance variations between the web browsers tested were significant and indicative of the potential web browser effect impact. The conclusions provide a warning for professional practice and increased awareness for potential loss and spoliation of evidence when investigating SNS.

REFERENCES

Chen, L., Xu, L., Yuan, X., & Shashidhar, N. (2015). “Digital forensics in social networks and the cloud: Process, approaches, methods, tools, and challenges”. Symposium conducted at the meeting of the 2015 International Conference on Computing, Networking and Communications (ICNC), 16-19 Feb. 2015.

Cheung, C. M. K., & Lee, M. K. O. (2010). A theoretical model of intentional social action in online social networks. *Decision Support Systems*, 49(1), 24-30.

Cohen, F. B. (2010). “Fundamentals of Digital Forensic Evidence”. In Handbook of Information and Communication Security, pp. 789-808. Berlin Heidelberg: Springer.

Dar, H., & Shah, A. (2013). “Analysis of SNS Popularity from Different Perspectives among Users”. Paper presented at the Information and Communication Technology for the Muslim World (ICT4M), 2013 5th International Conference on, 26-27 March.

Mumba, E. R. & Venter, H. S. (2014). “Testing and Evaluating the Harmonized Digital Forensic Investigation Process in Post Mortem Digital Investigations”. In Proceedings of the 2014 ADFSL Conference on Digital Forensics, Security and Law. Richmond, Virginia. May 28-29, 2014.

Jang, Y.-J., & Kwak, J. (2014). Digital forensics investigation methodology applicable for social network services. *Multimedia Tools and Applications*, 1-12.

Mingming, X. (2014). “Analysis of social networking services organizations' profit model based on Web2.0”. Symposium conducted at the meeting of the Service Systems and Service Management (ICSSSM), 2014.

Schneier, B. (2010). A Taxonomy of Social Networking Data. *Security & Privacy*, IEEE, 8(4), 88-88.

StatCounter. (2015). Top 5 Browsers to June 2015 | StatCounter Global Stats. Retrieved from http://gs.statcounter.com/#all-browser-ww-monthly-201506-201506-bar

Wang, Q., Woo, H., Quek, C., Yang, Y., & Liu, M. (2012). Using the Facebook group as a learning management system: An exploratory study. *British Journal of Educational Technology*, 43(3), 428-438.

Yue, G., Wang, L., Luan, H. & Chua, T. (2014). “Brand Data Gathering From Live Social Media Streams”. In Proceedings of International Conference on Multimedia Retrieval (ICMR '14). ACM, New York, NY, USA, pp. 169-177.

Zainudin, N., Merabti, M., & Llewellyn-Jones, D. (2010). “A digital forensic investigation model for online social networking”. In Proceedings of the 11th Annual Conference on the Convergence of Telecommunications, Networking & Broadcasting, Liverpool, pp. 21-22.

Zhang, H., Choudhury, M., & Grudin, J. (2014). “Creepy but inevitable?: the evolution of social networking”. In Proceedings of the 17th ACM conference on Computer supported cooperative work and social computing, Baltimore, Maryland, USA. Know your Enemy: Tracking Botnets. From http://www.honeynet.org/papers/bots/