Devious Chatbots - Interactive Malware with a Plot

Pan Juin Yang Jonathan, Chun Che Fung, and Kok Wai Wong

School of Information Technology, Murdoch University, South St, Murdoch Western Australia 6150
Jonathan.Pan.JY@gmail.com,{l.fung,k.wong}@murdoch.edu.au

Abstract. Many social robots in the forms of conversation agents or Chatbots have been put to practical use in recent years. Their typical roles are online help or acting as a cyber agent representing an organisation. However, there exists a new form of devious chatbots lurking in the Internet. It is effectively an interactive malware seeking to lure its prey not through vicious assault, but with seductive conversation. It talks to its prey through the same channel that is normally used for human-to-human communication. These devious chatbots are using social engineering to attack the uninformed and unprepared victims. This type of attacks is becoming more pervasive with the advent of Web 2.0. This survey paper presents results from a research on how this breed of devious Malware is spreading, and what could be done to stop it.

1 Introduction

Social robots in the forms of conversation agents or Chatbots have been put to practical use in recent years. Their typical applications are online help and as a cyber agent representing an organisation. However, one emerging trend is the development of devious chatbots with malicious intention. Effectively, they are malware in disguise. Malware, whose origin started as a biologically inspired innovation, is a malicious software treated by many as a detestable item and even something to be fearful of. The typical objective of Malware is to take control of the victims' PCs, steal information or cause more damage elsewhere in the form of Denial of Service attacks. In order to achieve the intent, many forms of attack vector are used by Malware deployers. This is a new form of Malware attack vector being used to increase its efficiency in delivering the attack on its target. Such Malware is going onto the collaboration platforms to launch their attacks by interacting with human. These Malware are appeared to be talking to the human! They are attempting to deceive the participant into believing that there is another human on the other end of the communication channel. If they are able to achieve this intent, will they be considered as having passed the Turing Test? Nicholas Carr tends to think they might be able to pass the test when the condition is right [1]. This survey paper reports the study done on this form of Malware and its attack vector. This paper also studies what is being done to counter this assault and provide a discussion on future research direction.

2 A New Online Infection

The Internet and the World Wide Web have provided many means to interact with one another. Email and instant messaging are among them. Popular Web 2.0 platforms like blogs, microblogs, social networking sites, social media sites and virtual worlds have further enriched the means to interact and communicate.

The popular forms of Malware are virus, worm, Trojan horse, bot and rootkit. A new form of Malware is lurking around seeking to devour "netizens" with greater sophistication and elusiveness. These Malwares are equipped with interactive capabilities to extend its reach, even onto the new online social platform. On 12 Dec 2007, it was reported that there is flirty chatroom bot that was attempting to steal identity information from those interacting with it [2]. Another is the Storm worm which caused a significant online disruption in 2007. Security researchers warned that up to
50 million computers were infected by that worm [3]. Another speculated that the Storm may have created the first cybercriminal controlled top 10 supercomputer by the sheer size of its botnet [30]. Its key characteristic was its cunning use of social engineering techniques as well as email and Web to propagate. In this study of Malware, it is limited to the social aspect of such robots and it is not considered interactive if the interaction is initiated by a human.

3 Interactive Malware

This study reports how Interactive Malware qualifies as a biologically inspired form of software with socially interactive capabilities, fulfilling certain aspects and characteristics of social robots. Also, this study covers the means in which the attack is being carried out and the form of deceptive interactivity used by this Malware to deceive its prey. Finally the various countermeasures used to fight them.

A. Biologically Inspired Software

The first Malware documented was by Cohen [32] as part of his PhD research. He demonstrated how computer viruses can reproduce themselves by taking advantage of the environment much like the biological viruses. A study by Kienzle and Elder [34] noted that the majority of the computer worms are derivative of worms found in nature. There are a number of notable similarities like infecting their host through an opening and replicating itself at the expense of the host. Both have the ability to spread autonomously without any human intervention. Both can be remain dormant for a period before striking. Both behaviours become more malignant when combining capabilities of other like entities. Researchers are looking to nature for ideas to counter such biologically inspired autonomous software. There are researches studying how immunological principles [36] and natural immune systems [37] can be used. The general consensus is that biologically inspired protection software would be part of everyone's desktop soon [35].

B. Malware using Collaboration Tools

As Malware is a malicious form of computer software, hence it would need to use the same electronic collaboration tools that we use to interact with us. Kickin is a Malware that uses email or internet relay chat (IRC) to interact and spread itself. It gathers email addresses stored within the infected PC's hard disk [4]. It then sends emails with different subjects and bodies via SMTP to its new targets. It also kills the antivirus and security processes to protect itself. The notorious Storm worm [3] sends themed or topical email messages. The 'TROJ_AGENT.ADB' Malware sends class reunion invitation emails with URL links to a site hosting a Trojan Malware [5]. According to Symantec [15], instant messaging (IM) is another popular channel used by such Malware. These Chatbots or Instant Messaging (IM) bots uses natural language dialogue systems used in gaming technologies to deceive its targets [6]. One such is the Russian developed CyberLover that can be found in chat-rooms and dating sites. Its intention is to gather identity information from its targets or lead them to websites containing Malware. Another chatty IM Malware is the Kelvir worm that uses predefined phrases to make small talk with the potential victims before sending a link to a malicious site [7].

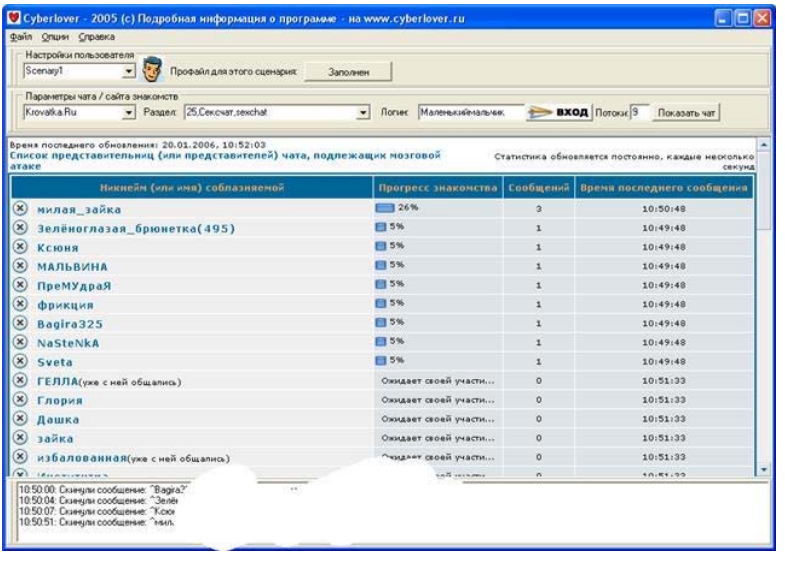

Fig. 1. CyberLover application snapshot (original picture is disfigured) [45]

Web 2.0 is also infested with Malware. Malicious bots have been known to have the capability to sign up and create blogging accounts on blogging sites like Blogger or Blogspot. Their intention of creating such blogs is to lure its targeted prey into its dens of Malware [8]. Social networking websites like Facebook has not been spared from the infectious Malware [9]. One such Malware is the Koobface worm [26]. Kaspersky Lab noted this worm and its variants infecting Facebook or MySpace accounts by sending a range of malicious commentaries or messages to friends' accounts with the intent to mislead them to websites containing Malware[10]. In

a position paper by the European Network and Information Security Agency [14], the term Malware 2.0 was coined for the new breed of Malware propagating in Web 2.0 space. According to security researchers from PandaLab [13], the future going forth will see Malware spreading actively among users in social networks. There is no interactive Malware noted in the virtual world and massive multiplayer online games, however Malware is pervasively used to attack such virtual resulting in theft of identity and virtual assets, extortion and terrorist attacks [27].

C. Socially Interactive Technology

According to the survey paper on socially interactive robots by Fong et al. [31], socially interactive robots (in this context, intelligent form of software) are robots that engages in peer-to-peer human-robot interaction with 'human social' characteristics like expressing emotions, communicate with high-level dialogue, establish social relationships, etc. These robots interact with humans through dialogues. A dialogue between a robot and human can only take place if there is a common symbol used. In this case, the symbol is natural language. An example of such intelligent software is by Goh and Fung [38] with their interactive human-like artificial intelligence Chatterbot called AINI (Artificial Intelligent Neural-network Identity). A study to use AINI to interact with humans via Instant Messenger showed that it did well in imitating human conversations and conversing with human-like artificial intelligence. AINI drew much interest and excitement from humans with its interactive capabilities [42]. Another Chatterbot named Natachata, written by a former rocket scientist Simon Luttrell, is used widely by porn chat merchants to provide mobile dirty talk through SMS text messages [39]. The customers here are led to think that they are communicating with young women or men. Chatterbot has been reported to engage in email exchanges. Epstein [40] cited how he was fooled into thinking that he was conversing with a Russian lady by the name of "Amélie Poulain". The conversation lasted months before he discovered he was conversing to a computer program. From the examples of Natachata and Amélie Poulain, there is another notable attribute in such socially interactive software robots. They have some forms of persuasive technology included. According to Fogg [43], such persuasive technologies can provide positive benefits to people. However they can also be used to achieve destructive purposes through manipulation and coercion of their victims. Researchers in socially adept technologies have found that we are generally not receptive towards such virtual peers. According to Angeli et al [41], one reason for this is the lack of common grounds between the human and virtual entity. However in the case of Malware, its social interactive capability may have some advantage as focuses on a specific common ground like lust to lure unwitting victims in. Malware developers now have the necessary technologies (natural language and persuasiveness) with common topics of interest to develop an effective socially interactive software robot. These robots or Malware could launch a social interactive form of attack on its unknowing targets. This form of social interactivity that such Malware will use is social engineering, popularly used by hackers and cybercriminals.

D. Malware using Social Engineering

According to SANS Institute [16], the phases involved in social engineering attack are Information Gathering, Developing Relationship with the targeted, Exploitation or manipulation of the targeted and finally Execution by getting the targeted to do the attacker's bidding. Malware uses the same social engineering phases to carry this form of attack vector. Information gathering may be initiated by the Malware or its deployer. Subsequently most Malware would attempt to establish a relationship with its targeted by finding common grounds with intention to lead quickly into the exploitation phase. This phase may involve manipulating the targeted to follow the provided URL to a website and subsequently execute the installation of the Malware. Consider Kickin Malware mentioned earlier [4]. In its attack, the information is gathered from the infected PC. Relationship is established through the randomized selection of email subjects and bodies with the hope of finding a common topic of interest with the targeted. Also given that the email originates from someone whom the targeted may possibly know, the relationship has been somewhat secured. The exploitation occurred by getting the recipient of the email to open the attached malicious file. Once the malicious package has been installed, the execution phase is completed.

Most of such Malware uses simple one way spam messages with no subsequent interaction involved. However more advanced Malware like CyberLover or Kelvir engages in greater extent of interactivity and may attribute themselves to be socially intelligent. The notable form of social engineering exploit used by such Malware is 'Likes and similarity' to seek a common ground with its targeted to establish trust quickly. This is one of the six human tendencies to social engineering according to National Infrastructure Security Coordination Centre (NISCC) from United Kingdom [17].

Malware have also been localized to cater to its prey of interest. Malware has been crafted for a specific country, language, organization and operating environment

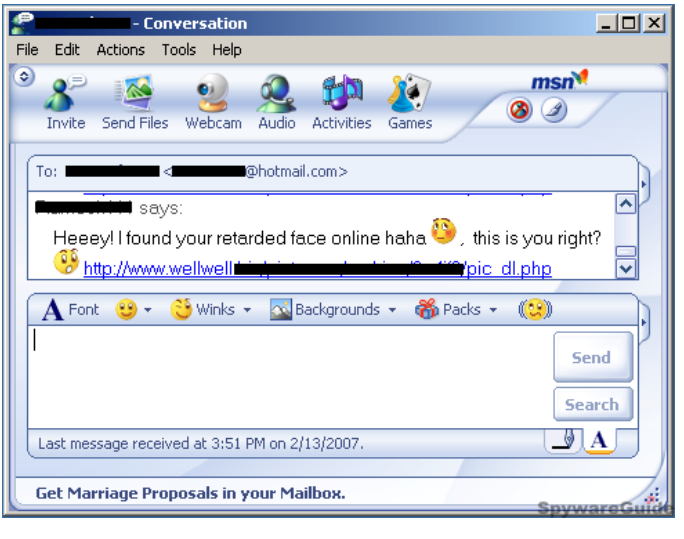

Fig. 2. Message sent through Messenger Window by Kelvir.EB Worm [44]

settings [18]. It has been adapted to communicate in Asian [18] and European languages [19]. A key enabler to this, as proposed by McAfee Avert Labs [20], is that Malware developers are not only skilled in computer programming but also in psychology and linguistics. Such custom built Malware would improve the outcome of relationship development phase.

Microsoft's Danseglio [22] commented that a notable reason for the success of social engineering attack is simply because of human's 'stupidity'.

E. Countermeasures

Anti-spam solutions may be used to filter malicious spam messages sent by Malware. However spam is still prevalent. Another way to counter such Interactive Malware is to prevent them from accessing the collaboration channels used by humans. This can be done at the point of registration to such channels. An attempt at this is the use of CAPTCHA (or "Completely Automated Public Turing test to tell Computers and Humans Apart") at registration that provides a challenge-response scheme with graphical representation of word(s) to stop automated software from successful registration. While such countermeasure had a good measure of success, it has been overcame by the advances in Malware technologies as well as gathering some human assistance. According to Websense [21], CAPTCHAs at popular websites like Google, Microsoft and Yahoo's web email services have been broken.

Researchers are studying how honeypots can be used to detect such assaults. Xie et al [11] is seeking to develop a honeypot for Instant Messaging called HoneyIM which uses decoy users to detect IM Malware. Another is a social honeypot by Webb et al that involves creating mock profiles in social networking communities and logs all communications made to these profiles and to automatically sift out deceptive spasms [12]. However there is one significant limitation noted with this solution. The honeypot's profiles prevents the establishment of any relational association with others in the social network. While this is understandable given dynamics of social networks, however Malware seeking to find new targets via existing relationships is not likely to interact with the honeypot's profiles.

Perhaps the best approach to counter such social engineering attacks is education and awareness [29]. This is especially useful when dealing with the weakest link in the battle against Malware. Governments, like the British government, are doing so [17]. SANS Institute had published papers on social engineering that could be used by defenders [16].

4 Future Research

According to Strickland [23], Web 1.0 is a library with lots of information available. Web 2.0 is about gathering of people to share information. Web 3.0 is like having a personal assistant who knows everything about oneself and helps one to gather the required information or invoke the required services. According to Wikipedia, Web 3.0 may usher in intelligent autonomous agents with natural language processing, machine learning and reasoning capabilities. Some measure of Artificial Intelligence has already been incorporated into Malware [28]. With such technology advancement, consider what Malware can do then. Hence fighters against Malware will want to monitor the development of this space closely as it provides a key enabler to Malware developers.

The typical classification of Malware like virus, worm and trojan horse are defined by its form of attack vector. Perhaps there should be one dedicated classification for such Malware. Also existing techniques in Malware analysis should be extended to identify interactive abilities

More research and development is required to protect our collaboration platforms from being used to launch attacks against us. Beyond seeking to improve measures to keep such Malware out, other measures can be developed to detect and stop malicious conversations.

5 Conclusion

Malware is encroaching into online lives in greater extent and begins to take on the disguise as social robots. Malware is seemingly able to communicate or interact with human. It may have created a digital 'mouth piece'. Chris Nuttal writes in Financial Times [24] that Web 2.0 creates 'a permissive society' where people share information freely, hence Malware will use this freely available and useful information to launch its attack. Malware is leveraging on popularly used collaboration platforms to interact with humans. However its ability to interact socially and intelligently is not as developed for most Malware using simple prescribed messages. However there is some advancement and a fare amount of research being done on socially interactive technology that Malware can leverage on. According to Thompson from BBC [25], he reckons that there is a real incentive for Malware developers to get the interactions done well so that its intended targets are more likely to be fooled into thinking that they are communicating with friends. Thompson also went further to suggest that such intelligent Malware could be used to find personal information, read emails and calendars on infected machines. Finally Thompson suggested that perhaps one day, such Malware may even pass Turing test and even win the Loebner Prize. It is urgently needed to manage this new form of Malware, perhaps should be better known as Malware 2.0, before it gains a deeper foot hold in the lives of netizens beyond the digital realm.

References

1. Carr, N.: Slutbot aces Turing Test*, December 8 (2007), http://www.roughtype.com/archives/2007/12/slutbot_passes.php (Accessed January 30, 2009)
2. Naughton, P.: Flirty Chat-Room 'Bot' Out to Steal Your Identity. December 12 (2007), http://www.foxnews.com/story/0,2933,316473,00.html (Accessed January 30, 2009)
3. Ironport, 2008 Internet Malware Trends – Storm and the Future of Social Engineering, Ironport (2008), http://www.ironport.com/malwaretrends/ (accessed January 30, 2009)
4. F-Secure, Virus Descriptions: Kickin, F-Secure, May 7 (2003), http://www.f-secure.com/v-descs/kickin.shtml (accessed January 30, 2009)
5. Baetiong, F.: 'Classmates Reunion' Used as Malware Ploy., Scientific American Mind, January 1 (2009), http://blog.trendmicro.com/classmates-reunion-used-as-malware-ploy/ (accessed January 30, 2009)

6. Rossi, S.: Beware the CyberLover that Steals Personal Data. Computerworld Australia, December 15 (2007), http://www.pcworld.com/printable/article/id,140507/printable.html (accessed January 30, 2009)
7. Schouwenberg, R.: Death of the IM-Worm? Viruslist.com, July 13 (2006), http://www.viruslist.com/en/analysis?pubid=191386185 (accessed January 30, 2009)
8. Websense, Google's 'Blogger' under attack by streamlined Anti-CAPTCHA operations for spam, Websense, April 24 (2008), http://securitylabs.websense.com/content/Blogs/3073.aspx (accessed January 30, 2009)
9. Helft, M.: Facebook Gets Friended by Malware, The New York Times, August 26 (2008), http://bits.blogs.nytimes.com/2008/08/26/facebook-gets-friended-by-malware/ (accessed Feburary 2, 2009)
10. Kaspersky Lab, Kaspersky Lab Detects New Worms Attacking MySpace and Facebook, Kaspersky Lab, July 31 (2008), http://www.kaspersky.com/news?id=207575670 (accessed Feburary 2, 2009)
11. Xie, M., Wu, Z., Wang, H.: HoneyIM: Fast Detection and Suppression of Instant Messaging Malware in Enterprise-like Networks. In: Twenty-Third Annual Computer Security Applications Conference (ACSAC 2007), pp. 64–73. Acsac (2007)
12. Webb, S., Caverlee, J., Pu, C.: Social Honeypots: Making Friends With A Spammer Near You. In: Sixth Conference on Email and Anti-Spam (2008)
13. PandaLab, PandaLabs' 2009 Predictions: Malware Will Increas. In: 2009, PandaLab, December 21 (2008), http://www.prweb.com/releases/2008/12/prweb1772314.htm (accessed Feburary 2, 2009)
14. European Network and Information Security Agency (ENISA), Position Paper – Web 2.0 Security and Privacy. European Network and Information Security Agency (2008)
15. Hindocha, N., Chien, E.: Malicious Threats and Vulnerabilities in Instant Messaging. Symantec Security response (2003)
16. Allen, M.: Social Engineering: A Means To Violate A Computer System. SANS Institute, InfoSec Reading Room (2006)
17. National Infrastructure Security Coordination Centre (NISCC), Social engineering against information systems: what is it and how do you protect yourself?, National Infrastructure Security Coordination Centre (NISCC), NISCC Briefing 08a/2006 (2006)
18. Nichols, S.: Malware gets up close and personal. IT News Australia, Feburary 22 (2008), http://www.itnews.com.au/News/70615,malware-gets-up-close-and-personal.aspx (accessed Feburary 2, 2009)
19. Dirro, T., Kollberg, D.: Malware Learns The Language. Sage, Thousand Oaks (2008)
20. McAfee Avert Labs, Localized Malware Takes Root. McAfee Avert Labs (2008)
21. Greenberg, A.: Robots In Disguise. Forbes.com, November 25 (2008), http://www.forbes.com/2008/11/25/cyber-security-bots-tech-identity08-cx_ag_1125cyberbots.html (accessed on Feburary 2 2009)
22. Naraine, R.: Microsoft Says Recovery from Malware Becoming Impossible. eWeek, April 4 (2006)
23. Strickland, J.: How Web 3.0 Will Work. HowStuffWorks, http://computer.howstuffworks.com/web-30.htm (accessed Feburary 4, 2009)
24. Nuttall, C.: The hidden flaws in Web 2.0., Global Technology Forum, Economist Intelligence Unit, The Economist, August 8 (2006), http://globaltechforum.eiu.com/index.asp?categoryid=&channelid=&doc_id=9168&layout=rich_story&search=footing (accessed Feburary 4, 2009)

25. Thompson, B.: Malicious worm that talks back. BBC News, December 12 (2005), http://news.bbc.co.uk/2/hi/technology/4520766.stm (accessed Feburary 4, 2009)
26. Finkle, J.: Destructive Koobface virus turns up on Facebook. Reuters, December 4 (2008), http://www.reuters.com/article/newsOne/idUSTRE4B37LV20081204 (accessed Feburary 4, 2009)
27. Muttik, I.: Securing Virtual Worlds Against Real Attacks. McAfee (2008)
28. Pan, J., Fung, C.C.: Artificial Intelligence in Malware – Cop or Culprit? In: The Ninth Postgraduate Electrical Engineering & Computing Symposium PEECS 2008. The University of Western Australia, Perth, Australia (2008)
29. Muncaster, P.: Firms must be alert to social engineering tricks. IT Week, September 26 (2007), http://www.vnunet.com/itweek/news/2199635/firms-alert-social-engineering (accessed Feburary 7, 2009)
30. Naraine, R.: Storm Worm botnet could be world's most powerful supercomputer. ZDNet, September 6 (2007), http://blogs.zdnet.com/security/?p=493 (accessed Feburary 9, 2009)
31. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots. Robotics and Autonomous Systems 42, 143–166 (2003)
32. Cohen, F.: Computer Viruses. PhD thesis, University of Southern California (1985)
33. Somayaji, A., Locasto, M., Feyereisl, J.: Panel: The Future of Biologically-Inspired Security: Is There Anything Left to Learn? In: The Proceedings of the 2007 New Security Paradigms Workshop
34. Kienzle, D., Elder, M.: Recent Worms: A Survey and Trends. In: WORM 2003, Washington, DC, USA, October 27 (2003)
35. Evans-Pughe, C.: Natural Defenses. Engineering & Technology (September 2006), http://www.theiet.org/engtechmag
36. Forrest, S., Hofmeyr, S.A., Somayaji, A.: Computer Immunology. Communications of the ACM 40(10) (October 1997)
37. Youansi, G.N.: Artificial Immune System. Communication and Operating Systems Group, Berlin University of Technology (2006)
38. Goh, O.S., Fung, C.C.: Intelligent Agent Technology in E-Commerce. In: Liu, J., Cheung, Y.-m., Yin, H. (eds.) IDEAL 2003. LNCS, vol. 2690. Springer, Heidelberg (2003)
39. Ward, M.: Has text-porn finally made computers human. BBC News, Feburary 20 (2004), http://news.bbc.co.uk/2/hi/uk_news/magazine/3503465.stm (accessed January 30, 2009)
40. Epstein, R.: From Russia, with Love. Scientific American Mind (2007)
41. Angeli, A., Johnson, G.I., Coventry, L.: The unfriendly user: exploring social reactions to chatterbots. In: International Conference on Affective Human Factors Design, Asean. Academic Press, London (2001)
42. Goh, O.S., Fung, C.C., Depickere, A., Wong, K.W.: An Analysis of Man-machine Interaction in Instant Messenger. Advances in Communication Systems and Electrical Engineering (2008)
43. Fogg, B.J.: Persuasive Technologies. Communications of the ACM 42(5) (May 1999)
44. SpywareGuide, Kelvir.EB. FaceTime Security Labs, http://www.spywareguide.com/product_show.php?id=3353 (accessed May 17, 2009)
45. Cyberlover, http://habrahabr.ru/blogs/cyberpunk/17263/ (accessed May 17, 2009)