Malware Biodiversity Using Static Analysis

Jeremy D. Seideman[1(✉)], Bilal Khan[2], and Antonio Cesar Vargas[3]

[1] The Graduate Center, City University of New York, New York, USA
jseideman@gradcenter.cuny.edu
[2] Department of Mathematics and Computer Science, John Jay College, CUNY, New York, USA
bkhan@jjay.cuny.edu
[3] NacoLabs Consulting, LLC, New York, USA
cesar@nacolabs.com

Abstract. Malware is constantly changing and is released very rapidly, necessarily to remain effective in the changing computer landscape. Some malware files can be related to each other; studies that indicate that malware samples are similar often base that determination on common behavior or code. Given, then, that new malware is often developed based on existing malware, we can see that some code fragments, behavior, and techniques may be influencing more development than others. We propose a method by which we can determine the extent that previously released malware is influencing the development of new malware. Our method allows us to examine the way that malware changes over time, allowing us to look at trends in the changing malware landscape. This method, which involves a historical study of malware, can then be extended to investigate specific behaviors or code fragments. Our method shows that, with respect to the method in which we compared malware samples, over 64 % of malware samples that we analyzed are contributing to the biodiversity of the malware ecosystem and influencing new malware development.

1 Introduction

When studying *malware*, it is tempting to treat it as artificial life [27]; whether or not malware can actually be considered artificial life is a study on its own. When studying artificial life, though, concepts from biology are used to aid in measurement and analysis [34]. Our study of malware requires us to adopt some of those concepts from biology. We begin to build our method with these concepts. To start, we consider each malware sample as an *organism*.

Minor differences between organisms do not necessarily separate them into different species – two organisms can exhibit some differences and still be considered the same species. With malware, this is seen with *variants* – malware that is based on earlier malware with small changes. Looking at variants leads to a discussion of what actually constitutes relationships among malware [11]. As malware writers often reuse other code or infection techniques used in earlier malware, we can say that the earlier malware has an influence on later "offspring" malware, in

R. Doss et al. (Eds.): FNSS 2015, CCIS 523, pp. 139–155, 2015.
DOI: 10.1007/978-3-319-19210-9_10

21. Newsome, J., Karp, B., Song, D.X.: Polygraph: automatically generating signatures for polymorphic worms. In: Proceedings of the 2005 IEEE Symposium on Security and Privacy, pp. 226–241 (2005). http://doi.ieeecomputersociety.org/10.1109/SP.2005.15
22. Annual Report Panda Labs - 2013 Summary (2013). http://press.pandasecurity.com/wp-content/uploads/2010/05/PandaLabs-Annual-Report_2013.pdf
23. Salthe, S.N.: Evolutionary Biology. Holt, Rinehart and Winston Inc., New York (1972)
24. Seewald, A.K.: Towards autmating malware classification and characterization. In: Proceedings of Sicherheit 2008, pp. 291–302 (2008). http://alex.seewald.at/files/2008-01.pdf
25. Seideman, J., Khan, B., Ben Brahim, G.: Determining vulnerability resolution time by examining malware proliferation rates. In: 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC), pp. 1678–1682 (2013)
26. Singhal, A.: Modern information retrieval: a brief overview. IEEE Data Eng. Bull. **24**(4), 35–43 (2001)
27. Spafford, E.H.: Computer viruses as artificial life. Artif. Life **1**(3), 249–265 (1994)
28. Threat explorer - spyware and adware, dialers, hack tools, hoaxes and other risks (2012). http://www.symantec.com/security_response/threatexplorer/
29. UPX: the Ultimate Packer for eXecutables - Homepage (2010). http://upx.sourceforge.net/
30. VirusTotal (2008). http://www.virustotal.com
31. VX heavens (2010). http://vxheaven.org/
32. Wagener, G., State, R., Dulaunoy, A.: Malware behaviour analysis. J. Comput. Virol. **4**(4), 279–287 (2008)
33. Wong, W., Stamp, M.: Hunting for metamorphic engines. J. Comput. Virol. **2**(3), 211–229 (2006). http://dx.doi.org/10.1007/s11416-006-0028-7
34. Woodberry, O.G., Korb, K.B., Nicholson, A.E.: Testing punctuated equilibrium theory using evolutionary activity statistics. In: Korb, K., Randall, M., Hendtlass, T. (eds.) ACAL 2009. LNCS, vol. 5865, pp. 86–95. Springer, Heidelberg (2009)