# Analysis and Detection of Evolutionary Malware: A Review

Pushpendra Dwivedi
Computer Science, FET, Rama University Uttar Pradesh, Kanpur

Hariom Sharan, PhD
Computer Science, FET, Rama University Uttar Pradesh, Kanpur

## ABSTRACT
Malwares now a days has become a big threat to the digital world around the globe. It target the network, system and penetrate into it, get access to the computers, brings down the servers, steal confidential information, ask for ransom, harm the critical infrastructure etc. to deal with the threats from these malwares and attack many anti- malwares have been developed so far. Some of them are based on the assumption that malwares do not change their structure. But with the with the advancements second generation malwares can create their variants that's why they are hard to detect. We present our survey on evolutionary malware and its detection techniques.

## Keywords
Malwares, Antimalware, Oligomorphic, Metamorphic, Polymorphic

## 1. INTRODUCTION
Internet has become more prevalent in our lives in shorter period than any other technology in the history. As popularity of cloud computing and internet of things (IoT) increasing, threats/attacks are also increasing. Laptop, smartphones, tablets, and other mobile devices have now become common everywhere due to their huge personal use and powerful features. User more frequently use internet on mobile devises, tablets etc., in these devises most popular operating system is Android. Threats in mobile app grow year to year.

According to Center for Internet Security report in the Top 10 Malware a significant increase in WannaCry, Emotet, Covter and ZeuS activity led to a 95% increase in top Top 10 activity [1].

A malware or malicious software is a program/code that is developed with the intent to enter into system without user authorization and takes undesirable actions such as damaging devices and stealing data. The term malware is too often used analogously with virus, even though the two are different. Malware is a concise, associated term used to refer viruses, worms, trojans, spyware, adware, rootkits, botnets etc. In today's digital world malwares are a big threat and are continuously evolving and growing with high complexity. The reason behind the increasing threat from malware is the wide spread use of Internet. An estimate shows that the web based attack increased 45% with over 4,500 new attacks each day, steal 75 records every second disrupting the victim in terms of confidentiality, integrity, availability of the user's data etc.

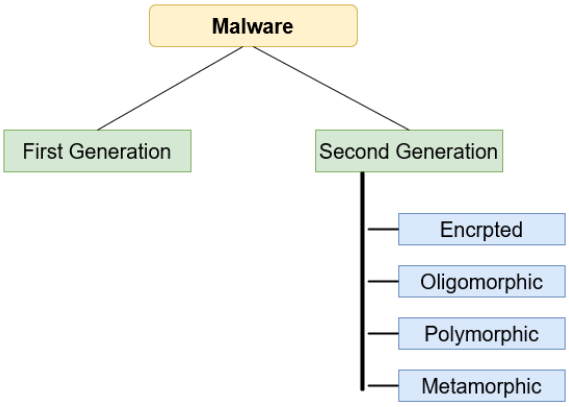


**Figure 1: Types of Malware**

## 2. MALWARES TYPE
Malwares are broadly classified into two categories: first generation and second generation. In first generation of malwares, internal structure of the malwares does not change. But in second generations of malware, the structure of malwares changes in every variant while their actions are maintained same. The second variant of malware is categorized into four categories: Encrypted, Oligomorphic, Polymorphic and Metamorphic Malwares.

### 2.1 Encrypted Malwares
Encryption was the first concealment techniques, and its ultimate aim is to change binary code of virus body is used for creating the 2nd generation malwares [3]. Generally Encrypted Malware consists of two parts; the encrypted body and a small piece of decryption code [2]. Each time it infects, encrypted malware automatically encodes itself differently, makes the body unique by using different key to hide the signature. CASCADE was the first encrypted virus [4]. Win95/Mad and Win95/Zombie uses the same technique were appeared in 32-bit Windows. The major motivation to use the encryption malware is to avoid static code analysis by encrypting the payload, make analysis process more difficult to delay the process of inspection, prevent tampering by producing man new variant and avoid detection [5].

Over past few years a steady increase is observed in the percentage of malware samples using TLS-based encryption to evade detection. Pattern matching is less effective in the presence of TLS sessions, leads to need of developing new methods that can accurately detect malware[6].

### 2.2 Oligomorphic Malware
The short comings of the encrypted malware was that there is a possibility to find the decryption mechanism led to the development of different decryptors and use them randomly while they are affecting other files. The decryptors are mutated

and changes their body. Initially in this type of malware, the developer encrypts the code that was capable of changing the decryptor lightly [5]. One way is to provide a set of different decryptors rather than one. W95/Memorial is capable of building 96 different decryptor patterns [7]. However, Oligomorphic malwares are not a good practices as they can be detected by signature based approach and tools [8].

## 2.3 Polymorphic Malwares

It contains two parts, one is the code decryptor to decrypt the second is body part but one part remains the same with each iteration. This leads to easier detection of the malware. During the execution of malware, mutation engine creates a new decryptor, adding varying length of NOP, permuting use registers, adding loops in the code joined with the encrypted malware body to construct a new variant of malware [9]. 1260 was the first known polymorphic malware that was written by Mark Washburn in 1990 [8].

## 2.4 Metamorphic Malwares

Metamorphic malwares are body-polymorphic with each iteration so that each succeeding version of code is different from the preceding one [8]. The technologies used by metamorphic malware is so sophisticated and complex i.e. Instead of generating new decryptor, a new instance (body) is created without changing its actions, the core functionality of the malware has to stay the same. Obfuscation techniques can be used to create new instances in metamorphic malware. The code changes makes it difficult for signature-based antivirus software programs to recognize. It is a challenging task to create a metamorphic malware without arbitrarily increasing the size. Only few malwares exhibit true metamorphic behavior [10], e.g. Phalcon/Skism Mass-Produced Code Generator, Second Generation virus generator, Mass Code Generator and Virus Creation Lab for Win32 were initially claimed to be metamorphic but were not. The first metamorphic virus was Win95/Regswap (in the year 1998) [12]. Win32/Ghost virus was created in 2000, with 3628800 different variants [12]. In 2001 W32/NGVCK was created with the help of Next Generation Virus Creation Kit (NGVCK), was one of the strongest metamorphic malware. Methods for metamorphic malware detection are described in [13].

# 3. MALWARE ANALYSIS AND DETECTION

Cyber threats have become ever-present pat in today's scenario and to combat the threat and attacks from the malwares, anti-malware softwares are developed. These softeares (anti-malwares) are commonly based on the assumption that the malware do not change their structure appreciably. Second generation malwares are hard to detect and the variant of second generation malwares are very much different to each other and threat from these malwares are increasing day by day. This creates a need that both researchers and anti-malwares developers should continually work to prevent these malwares from causing damage by the evolution of malwares. This section discus the various techniques used for the detection of malwares.

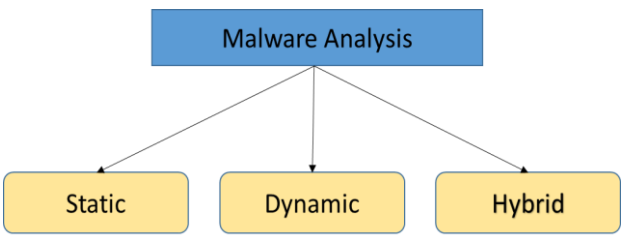


**Figure 2: Malware Analysis**

## 3.1 Signature based detection

Signature detection is the simplest approach which considers attack patterns as signatures and further compares signatures of known attacks to incoming attacks for detection [14]. Once the malware is identified by presence of malware infection/instance, unique sequences of bytes are extracted from it i.e. signature of the malware and identifies it by matching byte code pattern with the database of signature. This signatures are selected long enough to characterize a specific malware and this detection scheme is based on the assumption that malware can be characterized based in the signature. Signature based detection scans the file in to find the defined malware signature in the devices. Aho-Corasick algorithm scan for the exact matching [16]. They can be easily evaded with a slight mismatch or using obfuscation techniques like code re-ordering, no-ops. Signature based anti-malware are unable to detect unknown or even variants of known malwares [28]. Moreover it needs to update the malware definition for each variant that leads to exponential growth rate of signature database [28]. Yong Tang, Bin Xiao and Xicheng Lu, "Signature Tree Generation for Polymorphic Worms", IEEE TRANSACTIONS ON COMPUTERS,VOL. 60, NO. 4, APRIL 2011.

## 3.2 Heuristics based detection

Opposed to signature based detection heuristic approach uses rules and/or algorithm to look for instructions which may indicate malicious intent. New threats which exploded around and are continuing to emerge all the time. To counter this problem, the heuristic model was specifically designed to detect suspicious characteristics that may be found in unknown, new malwares and modified versions of existing as well as known malware. In this techniques, samples, in the suspicious file, program are disassembled to find a matching of the known malware pattern, if any. If the analysis result crosses the predefined threshold then the program is marked as infected [17].

The Heuristics method is a promising technique, however, it requires entire virtual environment to be installed, also it is prone to false alarm and if the malicious action is obfuscated successfully (e.g. within an encrypted file), it will evade detection. [19], which may make the system more vulnerable. To reduce the false alarm, we augment the results of detection techniques and combine it with another detection technique [20].

## 3.3 Machine Learning

Malwares now a days have multiple polymorphic layers or automatically change to a newer version to avoid detection. Heuristic based malware detection identifies a file by code fragments or hash code fragments. Malware detection with machine learning techniques is becoming popular. J. Z. Kolter et al. came with enhance decision tree working to produce better results [21]. M. R. Chouchane et al. proposed detection of malwares based on Hidden Markov Models [22].

Machine learning techniques not only detect known malwares but also act as knowledge base for the detection of new and variants of malware. The different machine learning techniques for the detection of malwares are Association Rule [23], Naive Bayes [26], Decision Tree [28], Data Mining [24], Neural Networks [26] and Hidden Markov Modes [12].

This technique act as an add-on feature with the standard detection methods. Generally, machine learning techniques are more computationally demanding then the standard anti-malware, hence it may not be suitable for end users. However, it can be implemented at enterprise gateway level to act as a

central anti-malware engine to supplement anti-malwares. Although, infrastructure requirement is costly, but it can help in protecting valuable enterprises data from the security threat and can prevent immense financial damages.

## 3.4 Malware Normalization

The malwares generated from advanced toolkits are difficult to detect [20]. Normalization techniques can be used to improve the detection rate of an existing anti-malware for the malwares that are generated from the toolkits (i.e. UPX and Mitsfall). Normalizer accepts the obfuscated version of malware and eliminates the obfuscation, after that the signature of the malware is extracted and compared with the signature of canonical form [25]. Christodorescu et. al. proposed a malware normalizer that handles code reordering, packing, and junk insertion [26] . Armor et. al., [27] proposed a generalized malware normalizer in the form of automata structures and use them for normalizing the metamorphic malwares. Recently a general malware normalizer has been proposed that can store multiple obfuscation methods for normalizing metamorphic malwares, which has a detection rate up to 81% [27].

# 4. CONCLUSION

Malware creators are a step ahead of anti-virus developers as there are many good software to create different variants of malwares. For detection of malwares different techniques such as: Heuristic approach, Signature based detection, Machine Learning and Normalization methods are used. Also there is no such technique exist that can detect zero day attack malware with 100% accuracy. Metamorphic and Polymorphic malwares uses obfuscation techniques like dead-code insertion, register reassignment, subroutine reordering, instruction substitution, code transposition and code integration are used to evade anti-malware scanners [29].

# 5. REFERENCES

[1] Symantec Corporation. 2012 Symantec Internet Security Threat Report, Symantec

[2] Rad, B., Masrom, M. and Ibrahim, S. “Camouflage in Malware: From Encryption to Metamorphism”, International Journal of Computer Science and Network Security, 2012, 12: 74-83.

[3] Beaucamps, P. “Advanced polymorphic techniques International Journal of Computer Science”, 2001, 25: 400411.

[4] Rad, B. B., Masrom, M. and Ibrahim, S. “Evolution of Computer Virus Concealment and Anti-Virus Techniques: A Short Survey, International Journal of Computer Science Issues, 2011, 8:113121.

[5] Shah, A. Approximate Disassembly using Dynamic Programming [PhD. Thesis], San Jose State University, US, 2010.

[6] B. Anderson and D. McGrew. Identifying Encrypted Malware traffic with Contextual Flow Data. In Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security, AISec ’16, pages 35-46, 2016.

[7] Cho, Y. and Mangione-Smith, W. High-performance context-free parser for polymorphic malware detection, United States Patent US 2006113722, 2009 April 18.

[8] Austin, T. H., Filiol, E., Josse, and Stamp, S. M. “Exploring Hidden Markov Models for Virus Analysis: A Semantic Approach, Proceedings of the 46th Hawaii International Conference on System Sciences, Wailea, HI, USA, 2013, Jan 7-10, 50395048.

[9] Ferrie, P., Corporation, S. and Monica, S . “HUNTING FOR METAMORPHIC”, Proceedings of the Virus Bulletin Conference 2001, Czech Republic, Prague, 2001 Sep 27-28, 123144.

[10] Ddcreateur, ANTIVIRUS 2004, [Database on the Internet]. Codes-sources library. [updated 2004 March 26; cited 2013 Oct 1]. Available from http://files.Codessources.com/fichierfullscreen.aspx?id=21418&f=virus signatures.txt&lang=en

[11] Tran, N. and Lee, M. “High performance string matching for security applications”, Proceedings of the International Conference on ICT for Smart Society, Jakarta 2013 June 13-14, 15. 11

[12] Mathur, K. and Hiranwai, S. “A Survey on Techniques in Detection and Analyzing Malware Executables”. International Journal of Advanced Research in Computer Science and Software Engineering, 2013, 3: 422428.

[13] E. Konstantinou, “Metamorphic virus: Analysis and detection,” 2008, Technical Report RHUL-MA-2008-2, Search Security Award M.Sc.thesis, 93 pages

[14] Harley, D. and Lee, A. “Heuristic AnalysisDetecting Unknown Viruses”, [White paper] Eset, 2007, [cited 2013 Oct 1]. Available from http://www.eset.Com /us/resources/white-papers/Heu- ristic Analysis.pdf

[15] Wong, W. and Stamp, M. “Hunting for metamorphic engines”, Journal in Computer Virology, 2006, 2: 211229.

[16] Mitchell, T. M. “Machine learning”, Burr Ridge, IL: McGraw Hill, 1997.

[17] Moskovitch, R., Yuval, E. and Lior, R. “Detection of unknown computer worms based on behavioral classification of the host”, Computational Statistics & Data Analysis, 2008, 52: 4544-4566.

[18] Alazab, M. and Venkatraman, S. “Zero-day malware detection based on supervised learning algorithms of api call signatures”, Proceedings of the Ninth Australasian Data Mining Conference, Ballarat, Australia 2011 Nov, 121: 171182.

[19] Moskovitch, R., Elovici, Y. and Rokach, L. “Detection of unknown computer worms based on behavioral classification of the host, Computational Statistics & Data Analysis, 2008, 52:45444566.

[20] Siddiqui, M., Wang, M. C. and Lee, J. “A survey of data mining techniques for malware detection using file features, Proceedings of the 46th Annual Southeast Regional Conference, New York, USA, 2008, March 28-28, 509-510.

[21] J. Z. Kolter and M. A. Maloof, “Learning to detect and classify malicious executables in the wild,” Journal of Machine Learning Research, vol. 7, pp. 2721–2744, December 2006, special Issue on Machine Learning in Computer Security.

[22] M. R. Chouchane, A. Walenstein, and A. Lakhotia, “Using Markov Chains to filter machine-morphed variants of malicious programs,” in Malicious and Unwanted Software, 2008. MALWARE 2008. 3rd International Conference on, 2008, pp. 77–84

[23] Y. Ye, D. Wang, T. Li, and D. Ye, “Imds: intelligent malware detection system,” in KDD, P. Berkhin, R.

Caruana, and X. Wu, Eds. ACM, 2007, pp. 1043–1047.

[24] Xu, M., Wu, L., Qi, S., Xu, J., Zhang, H., Ren, Y. and Zheng, N. "A similarity metric method of obfuscated malware using function-call graph", Journal of Computer Virology and Hacking Techniques. 2013, 9:3547. 12

[25] Xu, J., Sung, A. H., Chavez, P. and Mukkamala, S. "Polymorphic malicious executable scanner by API sequence analysis", Proceedings of the Fourth International Conference on Hybrid Intelli- gent Systems, Kitakyushu, Japan, 2004, Dec 5-8, 378-383.

[26] Christodorescu, M., Johannes, K., Jha, S., Katzenbeisser, S. and Veith, H. "Malware Normalization", University of Wisconsin, Madison, Wisconsin, USA, 2005 November. Report No: 1539.

[27] Armoun, S. E. and Hashemi, S. "A General Paradigm for Normalizing Metamorphic Malwares, Proceedings of the 10th International Conference on Frontiers of Information Technology, Islamabad, Pakistan, 2012, Dec 17-19, 348353.

[28] Yong Tang, Bin Xiao and Xicheng Lu, "Signature Tree Generation for Polymorphic Worms", IEEE TRANSACTIONS ON COMPUTERS,VOL. 60, NO. 4, APRIL 201.

[29] You, I. and Yim, K. "Malware Obfuscation Techniques: A Brief Survey, Proceedings of IEEE International Conference on Broadband, Wireless Computing, Communication and Applications, Fukuoka, 2010, Nov 4-6, 297300.