American Journal of Interdisciplinary Research and Development
ISSN Online: 2771-8948
Website: www.ajird.journalspark.org
Volume 02, March, 2022

# MALWARE ANALYSIS NETWORK IN UZBEKISTAN (MIT)

Husanboy Shoraimov Uktamboyevich
(Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi, Assiatant Teacher of the Department "Systematic and Practical Programming", khusan@shoraimov.uz, +998901244448)

**Abstract**
Malware is one of the problems really existing in the modern post-industrial society. Hackers continuously develop novel techniques to intrude into computer systems for various reasons, so many security researchers should analyze and track new malicious program to protect sensitive information for the computer system. In this paper, we integrate the Interval Type-2 Fuzzy Logic System (IT2FLS) with malware behavioral analysis: Malware Analysis Network in Uzbekistan (MAN in Uzbekistan, MiT). The core techniques of MiT are as follows: (1) automatically collect the logs the difference operation system to extract unknown behavior information.

**Keywords:** Type-2 Fuzzy Sets, Interval Type-2 Fuzzy Logic System, Web Ontology Language, Malware Behavioral Analysis, MiT.

## INTRODUCTION

In recent years, malicious software (malware) and the common infection vectors that the malware use to reach end users are important threat and root cause of many security problems on the Internet. The malware tsunami is overwhelming for security researchers across the world. How to reduce the damage caused by malware and how could a malware evade anti-virus software also are important issue for governments, universities, commercial organizations, and so on [1, 2]. Also, detecting and analyzing them become harder and harder, especially because the malware are getting more complicated. Many security researchers have proposed some new defenses to protect user the confidential data for the computer system. Unfortunately, the hackers always one step ahead with security researchers that exploits a previously unknown vulnerability in a computer system which causes computer systems to be damaged and confidential data to be stolen [3]. The heuristic and signature detection technologies which are two popular approaches to malware analysis for a few years. In order to rapidly defend against unknown malicious attack, many security researchers and traditional malware detection systems use the signature matching techniques to develop an automatic effective analysis tool for detecting malware. However, this approach can be easily circumvented the attack of the malware because the polymorphic characters or metamorphic features of malware will mutate their

signatures when malicious software is spread from one host to another one [4]. Indeed, many researches provide malware analysis for monitoring malware's actions while it is running under a controlled environment like Virtual Machine (VM). This approach is so-called Virtual Machine Monitor (VMM), which can identify the malware behavior and what the malware has modified in the file system or the registry to quickly recover from the malware infection state. Therefore, a VMM approach is suitable for the malware analysis, and most malware analyses are carried out under VM.

There are many areas including modeling, control, and data mining which have been successfully applied Type-1 Fuzzy Set (T1FS) and Type-1 Fuzzy Logic System (T1FLS) [7, 8]. However, different experts construct different membership functions to represent the same object. A type-2 fuzzy set is characterized by a fuzzy membership function (MF), i.e. the membership value (or membership grade) for each element of this set is a fuzzy set in [0, 1], unlike a T1FS where the membership grade is a crisp number in [0, 1]. T2FS also has been widely developed and successfully used in many practical real-world applications and express more fuzzy semantics of humans' thoughts, including signal processing, human silhouette extraction, diet application, and pattern recognition design. Interval Type-2 Fuzzy Set (IT2FS) is a special cases of T2FS, and it is currently the most widely used because of the reduction of computational cost .

Ontologies play a key role in the Semantic Web, however it has been widely pointed out that the traditional ontology is not suitable to deal with uncertain, vague, and imprecise knowledge to characterize the real-world scenarios. As a consequence, there is a growing interest in fuzzy ontologies, which combine ontologies and fuzzy logic theory. Therefore, the fuzzy ontology is emerging as a useful methodology for knowledge representation in several semantic-oriented applications. This paper tries to integrate then abovementioned different kinds of the approaches to solve the uncertain problem with the cyber security.

## II. BACKGROUND KNOWLEDGE

A. Malware Behavioral Analysis Overview Internet and personal computers have rapidly advanced [1] in recent years so hackers and their malicious software packages which are a computer program which has destructive purposes such as Botnet, Virus, Backdoor, and Trojan, attempt to steal user' data or illegally control computer systems. Growing the Internet significantly increases the variety and complexity of malware; hence, malware poses serious security effects on information societies. When the anti-virus software or traditional detection approach can't detect or defense those virus attacks, it can cause the system broken, network paralysis or heavy loss. Therefore, security researchers are always proposing some new defenses to protect users' personal, valuable, and confidential information. However, they always fall behind the hackers. In other words, the battle between hackers and security researchers never has

an ending [3]. In order to rapidly defend against unknown malicious attack, many security researchers and traditional malware detection systems use the signature matching techniques to develop an automatic effective analysis tool for detecting malware. However, this approach can't detect an attack from unknown malware or its variant and can be easily circumvented the attack of the malware because two interesting kinds of malware which attempt to be hidden from detectors' signature are polymorphic and metamorphic malware. The polymorphic characters or metamorphic features of malware will mutate their signatures when the malicious software is spread from one host to another one [3]. Obviously, it is so hard to detect this type of malware by classic signature base method. Nevertheless, in the earlier times, the most effective and accessible method for capturing malware was signature-based detection and it is a popular approach for the malware analysis [5] and another limitation of most existing detection techniques is that they are good at detecting only known forms of malware. They tend to be ineffective when faced with new forms of malware. To resolve the main drawback of classic signature based method, which is an inability in detecting unknown malware, behavior based (or heuristic) solutions are developed. Behavior-based malware detection approach attempt to find the pattern of malware behavior for further recognitions of similar malicious behavior and which has a greater potential for identifying previous unknown malware [4]. They are able to fulfil the malicious activity detection during code execution by trying to trace any suspicious.

A VMM approach is suitable for the malware analysis, and most malware analysis is carried out under virtual machines. However, the transparency of the majority of VMs that are designed to detect the malware is not well enough until now. Malware developers have noticed such a situation that they have developed several techniques such as Anti-VM techniques to detect whether the malware is running under a virtualized environment or not. With the Anti-VM techniques, this causes the hackers to easily find the solutions to detect if the developed malware is running under VM-based environment and then avoid the detection from VMM. In most cases, malware can easily escape from the detection of the VMM to block the behavior of the propagation so that the detected malicious behavior from VM-based malware analysis sometimes may be different from the results of the physical environment.

American Journal of Interdisciplinary Research and Development
ISSN Online: 2771-8948
Website: www.ajird.journalspark.org
Volume 02, March, 2022

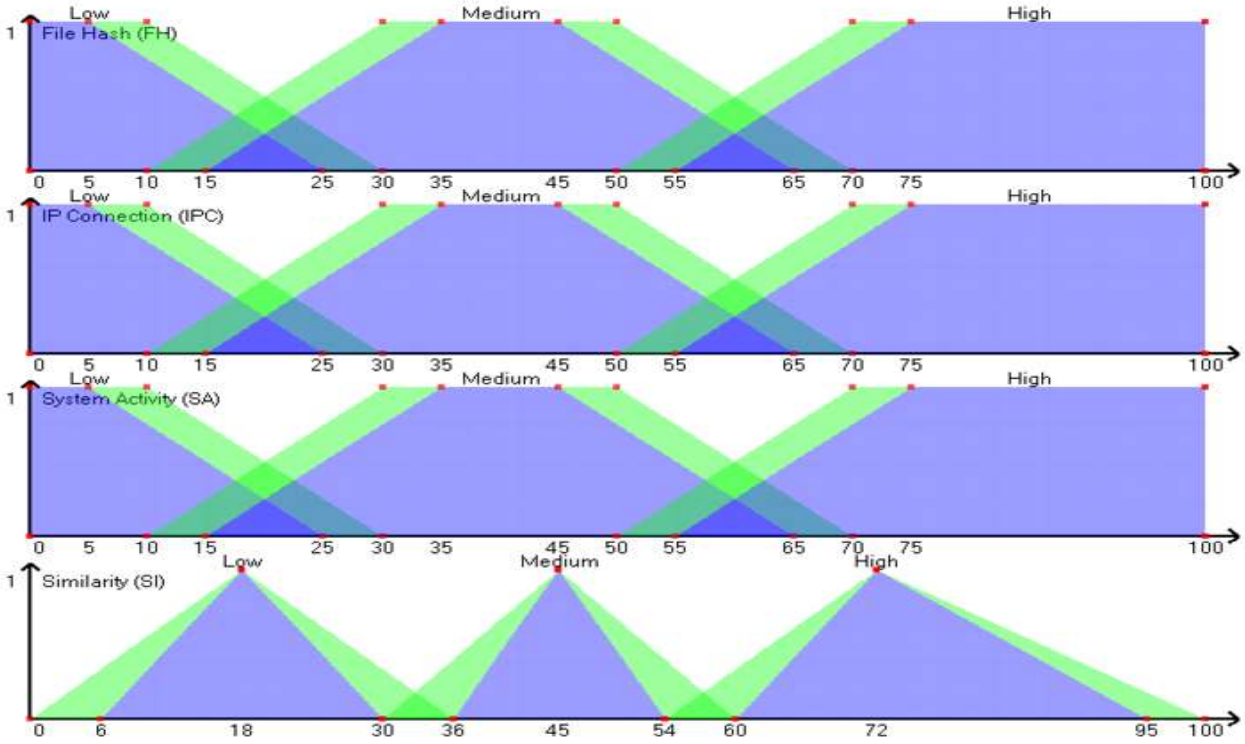


*Figure-1: Type-2 fuzzy sets for the type-2 fuzzy variables: File Hash, IP Connection, System Activity, and Similarity.*

B. Fuzzy Ontology Model Overview Typically, joint exploitation of the fuzzy ontologies to be one supported framework for designing the fuzzy inference systems is one of the key research topics in the soft computing research areas. There are many researchers exploring the use of the fuzzy ontologies, for example, Lee et al. [4] proposed a fuzzy ontology for designing an intelligent decision making system for summarization system. Quan et al. presented the automatic fuzzy ontology generation for semantic help desk support and the automatic fuzzy ontology generation for semantic web. The fuzzy ontology model is introduced in this section. In order to make both machine and human to understand the designed ontology, Web Ontology Language (OWL) is used in this paper to express the built ontology. In addition, we use Protégé to generate OWL for constructing the knowledge base of the ontology. Table I shows the built four-layer type-2 fuzzy ontology model by two views, and the brief descriptions between these two views including machine understandability and human semantic understandability. It enables developers to share common concepts and terms, and allows them to be described in a simple language. Most works are based on clustering algorithms aiming to identify similar behavioral patterns of malwares derived from network traffic traces recorded from communication and potential malicious activities such as scanning, spamming, binary downloading, and exploit attempts. Obviously, it is so hard to detect this type of malware by classic signature base method. Also, most malware behavioral analysis toolkits still need domain experts to interpret the

important semantics for the detected information of the malicious behavior and then judge it is a malware or not. We focus on recognition of unknown before malware which can't be detected by using traditional signature or rule-based detection techniques, oriented on search for concrete malware samples and families. Therefore, this paper tries to exploit an ontological view of the malware behavior to define a more general and efficient detection methodology. Ontology provides a means to clarify the concepts and semantics of the malware to avoid from some conceptual confusion. Additionally, ontology can share common concepts or relationships to allow the problems of the malware analysis to be described in a formal semantic platform among intelligent agents or malware behavioral analysis toolkits.

## III. INTELLIGENT HYBRID MALWARE ANALYSIS MECHANISM

A. Malware Analysis Network in Uzbekistan (MiT) Dynamic analysis of malware has received a lot of attention in the research community. There are many analysis systems such as cuckoo (http://www.cuckoosandbox.org), CWSandbox , and Anubis , use Virtual Machine (VM) systems to monitor malware behaviors. This is because one challenge in malware analysis involves collecting useful data without risking experimenters' machines or systems. VM technology also has many advantages (light, fast etc...) however, there are many Anti-VM techniques which are used to hinder the collection, analysis, and reverse engineering features of the VM based malware analysis platform. Therefore, malware researchers may receive inaccurate analysis results from VM based malware analysis platforms.

In this paper, based on our previous physical environment analysis toolkit: TWMAN , we re-develop and then propose a new generation toolkit to analyze malware behavior (Malware Analysis Network in Uzbekistan, MiT; also known as MAN in Uzbekistan) to resolve some weaknesses of TWMAN. Figure 1 shows the screenshot of its option on MiT boot, and Figure 1 shows the system structure and workflow of the MiT, We use three items to describe the improvements in MiT as follows:

- Re-design and mash up a VM as an analyzed platform to be the distributed structure to decrease the hardware cost, also pre-check hash value by ssdeep to improve the weakness.
- Use toolkits to monitor the important directories of the client and save the time that the system's image stores back to the server.
- Implement the proposed IT2FLS to identify the malware behavior.

Therefore, MiT is a virtual-physical hybrid environment and has been developed to automate malware behavior analysis, then to detect the unknown malware based on known malware, and finally to synchronize the analyzed reports and malware samples for all users to resolve the above-mentioned troubles.

*B. IT2FS-based Malware Behavior Knowledge Base for MiT* OWL enables a suitable representation of malware knowledge, it is not able to apply the advanced inference mechanism to derive the additional imprecise and vague knowledge in the scenario of

the detection of the malwares. Indeed, ontology also includes a vocabulary of terms and the specifications of the terms' meanings to express the relations among concepts and definitions.

In this paper, there are three input type-2 fuzzy variables and one output type-2 fuzzy variable defined in *MiT*. We define three linguistic terms, including Low, Median and High for the input fuzzy variables *File_Hash (FH)*, *IP_Connection (IPC)*, and *System Activity (SA)*, respectively. Additionally, the output type-2 fuzzy variable *Similarity (SI)* also contains three linguistic terms, including Low, Median and High utilized in this paper.

Precisely, *FH* is a malware information which is computed by the ssdeep toolkit, and denotes a hash value bounded in an interval [0, 100] to express the similar level to a known malicious sample. *IPC*, ranging between 0 and 100, denotes the counted number of TCP/IP connections from InetSim (http://www.inetsim.org) to express the similar level to a known malicious sample calculated by the regular expression. *SA* denotes the generated behavioral similarity between the analyzed malicious sample and known malicious sample which is calculated by the regular expression and ranges from 0 to 100. _ Instance layer contains *Similarity* (*SI*), the type-1 fuzzy set layer, and the type-2 fuzzy set layer. *SI* calculates the similarity between an unknown malware and a known malware according to the values of the *FH*, *IPC*, and *SA*, which come from PDF documents, DLL files, Windows Executables, and Office Documents existing in Microsoft Windows 7, Microsoft Windows XP and so on.

## IV. SIMULATION RESULTS

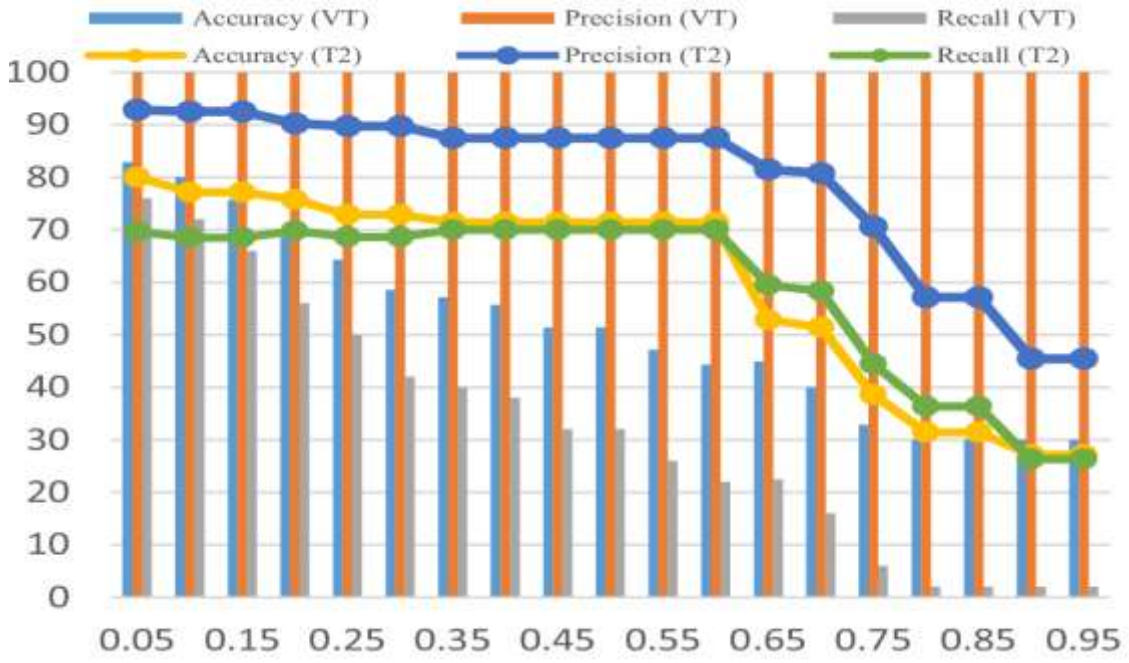


*Figure 2. Accuracy, precision, and recall curves when using VT and IT2FS to make a analysis.*

One of the novel attacking models by emails on the Internet is a very serious security problem for the computer system so-called as Advanced Persistent Threat (APT) until now. Therefore, this paper try to reduce a complex task of analyzing a huge amount of malware for the e-mails to establish a knowledge model for future analysis work. Based on MiT, we partnered with Acer eDC company in Taiwan to produce a scanner for the e-mail attachments, then analyze if there exits malware, and finally generate the reports. In this paper, we first download the 1360 known malicious samples from malwaretipss (http://malwaretips.com) and construct the established physical-virtual hybrid environment for testing the proposed approach. Second, the collected 1360 known malicious samples are used as the compared baseline. Third, 50 known malicious samples provided by Acer eDC company and additional 20 known non-malicious samples generated by ourselves are used as the experimental samples for the proposed IT2FLS.

The performance of the proposed approach is evaluated according to the criteria such as accuracy, precision, and recall. Figure 1 shows the curves of accuracy, precision and recall when we use the VT website to simulate the 70 experimental samples. All values of precision are 100% for each threshold in Figure 1. The reason is because no any Antivirus vendors on the VT website analyze 20 known non-malicious experimental samples to be a malware. Besides, Figure 1 also shows that accuracy and recall has a tendency to decrease when the threshold value is increased. The curves of accuracy, precision and recall for using the IT2FLS to analyze the 70 experimental samples based on the 1360 known malicious samples are also shown in Figure 2. Most values of accuracy in Figure 2 are higher than the ones in Figure 2 when the threshold is higher than 0.5. However, if a brand new malware is uploaded to the VT website, the probability that any Antivirus vendor judges it is a malware is relatively very low because these Antivirus vendors have no its signatures. On the other hand, users cannot know if the attachment contains the malware or not until they manually upload it to the VT website to make the analysis. After that, VT website still cannot give users an answer because VT only tells the users how many Antivirus vendors consider it a malware. Hence, *MiT* with the proposed IT2FLS has some strengths to improve the above weaknesses. Its strengths are as follows: (1) *MiT* is able to automatically proceed a malicious analysis; (2) Current malware-analyzing toolkits on the market only can do the analysis but cannot give users an answer after analyzing a suspicious unknown file or attachment. On the contrary, *MiT* can give users a possibility that the analyzed file or attachment contains a malware; (3) *MiT* can do the malicious analysis no matter whether the malware is with Anti-VM techniques because *MiT* is capable of operating in a virtual-physical hybrid environment; (4) *MiT* can simultaneously proceed the malicious analysis on various operation systems to reduce the probability of making an incorrect judgment only when the malware is actuated under a specific environment.

## CONCLUSION

In this paper, we present a novel interval type-2 fuzzy ontology methodology for an automation hybrid malware analysis system to analyze malware behavior. Analyzing the malware behavior is full of uncertainty, the problem of detaching the similarity behavior from the known malicious behavior to be the baseline becomes even more complicated. Compared to the results running on VT website, the simulation results also show similar results for the malicious detection. In other words, by utilizing the IT2FLS, the proposed system obtains the good results for unknown and uncertain malware's behavioral extraction and analysis. The experimental results also show that the proposed IT2FLS can perform effectively. In the future, we will continue analyzing the behavior of the known malicious samples and define more reasonable range for the T2FS of the fuzzy variable with machine learning and big data analysis approach to improve the proposed approach's performance. This also is extremely challenging future work and essential for a security solution to be useful in the everyday computing world.

## REFERENCES

1. H. D. Huang, G. Acampora, V. Loia, C. S. Lee, and H. Y. Kao, "Applying FML and Fuzzy Ontologies to Malware Behavioral Analysis,"
2. S. Y. Dai, Y. Fyodor, S. Y. Kuo, M. W. Wu, and Y. Huang, "Malware Profiler Based on Innovative Behavior-Awareness Technique," .
3. S. Y. Dai, Y. Fyodor, M. W. Wu, Y. Huang, and S. Y. Kuo, "Holography: a behavior-based profiler for malware analysis,".
4. G. Wagener, R. State, and A. Dulaunoy, "Malware behaviour analysis," Journal in Computer Virology, vol. 4, no. 4, pp. 279-287, 2008.
5. M. K. Sun, M. J. Lin, M. Chang, C. S. Laih, and H. T. Lin, "Malware Virtualization-Resistant Behavior Detection," .
6. C. S. Lee, Z. W. Jian, and L. K. Huang, "A fuzzy ontology and its application to news summarization,".
7. G. Acampora and V. Loia, "Fuzzy control interoperability and scalability for adaptive domotic framework,".
8. H. Hagras and C. Wagner, "Towards the Widespread Use of Type-2 Fuzzy Logic Systems in Real World Applications".
9. D. Wu, "On the Fundamental Differences Between Type-1 and Interval Type-2 Fuzzy Logic Controllers,".
10. C. S. Lee, M. H. Wang, and H. Hagras, "A Type-2 Fuzzy Ontology and Its Application to Personal Diabetic-Diet Recommendation," .
11. U. Bayer, I. Habibi, D. Balzarotti, E. Kirda, and C. Kruegel, "Insights Into Current Malware Behavior," 2nd USENIX Workshop on LargeScale Exploits and Emergent Threats (LEET), Boston, MA, 2018.

12. Akhmedovna Y. T. To develop students' knowledge, skills and competencies in the organizational and technical aspects of essay //ACADEMICIA: AN INTERNATIONAL MULTIDISCIPLINARY RESEARCH JOURNAL. – 2021. – Т. 11. – №. 2. – С. 914-918.