

Tutorial
on
AI BASED MALWARE DETECTION
by

Mohit Sewak
Principal Applied Researcher
Microsoft Security Research (MSecR)
Microsoft R&D, India
`mohit.sewak@microsoft.com`
and
Hemant Rathore
Assistant Professor
Dept. of CS & IS
BITS Pilani, India
`hemantr@goa.bits-pilani.ac.in`

1 Abstract:

Often computer/mobile users call everything that disturbs/corrupts their system a *VIRUS* without being aware of what it means or accomplishes. This tutorial systematically introduces the different malware varieties, their distinctive properties, different methods of analyzing the malware, and their detection techniques.

2 Introduction:

Today computing devices like laptops, mobile phones, smart devices, etc., have penetrated very deep into our modern society and have become an integral part of our daily lives. Currently, more than half of the world's population uses computers/mobile devices for their professional/personal needs. However, these computing devices are targeted by malware designers encouraged by profits/gains associated with the attack. According to a recent report, monetary losses due to cybercrime are expected to reach 10 trillion dollars annually by 2025. The primary role in providing defense against malware attacks is designed and developed by the anti-malware community (researchers and the anti-virus industry). Traditionally anti-viruses are based on the signature, heuristic, and behavior based detection engines. However, these engines are unable to detect next-generation polymorphic and metamorphic malware. Thus researchers have started developing malware detection engines based on machine learning to complement the existing anti-virus engines. However, there are many open research challenges in these models like adversarial robustness, explainability, fairness, etc., which we are going to discuss in detail during the workshop.

3 Objectives:

This workshop will cover fundamental techniques, limitations, open research problems and future directions in the field of malware analysis and detection. Following are the three specific learning outcomes:

1. Audiences will get familiarity with different types of malware and their detection techniques.
2. Applications of classification and clustering based frameworks for malware detection.

3. Overview of significant research problems in the area of malware analysis and detection, results, and conclusions from the recent research papers.

4 Target Audiences:

Senior undergraduate students (B.E.), postgraduate students (M.E./M.Tech./M.S.), PhD. students, faculty members, and researchers working or interested in the area of malware analysis and detection.

5 Duration:

3 hours

6 Pre-requisite

1. Basic knowledge of the operating system (Windows/Android)
2. Understanding of assembly codes & C programming language
3. Familiarity with classification and clustering techniques (Desirable)

7 Topics to be covered:

1. Introduction to Malware
2. A short history of Malware (virus to malware)
3. 1st Generation Malware
4. 2nd Generation Malware
5. Traditional Malware Detection Systems
6. Static Malware Analysis
7. Challenges in Static Analysis
8. Dynamic Malware Analysis
9. Challenges in Dynamic Analysis
10. Malware Detection as a Classification Problem
11. Challenges in AI based Malware Detection Systems
 - Adversarial Robustness in Malware Detection Systems
 - Explainability in Malware Detection Systems
 - Fairness in Malware Detection Systems
 - Data Challenges and Routing based Malware Classification
12. Alternate Mechanisms for Malware Detection
 - Advanced Metamorphic Malware Generation
 - Malware Normalization as a Defense Strategy
13. Open Research Problems and Future Directions

8 Hands-on Session:

Laboratory activities will involve analyzing and handling malicious code on a test system. Virtual machines can be used but it is not recommended to use the organization's laptop in laboratory activity.

9 More Details:

1. The tutorial is designed and delivered by specialists from industry and academics.
2. The tutorial will also cover a hands-on session for practical engagement.
3. The last three tutorials had an excellent response with more than 100 participants in each tutorial.

10 Previous edition of the Tutorial @ other Conferences:

1. Malware Detection @ *AIMLSystems 2022* by Hemant Rathore and Mohit Sewak
2. Malware Detection @ *IEEE Future Networks World Forum 2022* by Hemant Rathore and Ashu Sharma
3. Malware Analysis and Detection @ *IEEE BlackSeaCom 2021* by Hemant Rathore and Ashu Sharma
4. Malware Analysis and Detection @ *IEEE CAMAD 2021* by Hemant Rathore and Ashu Sharma
5. Malware Analysis and Detection @ *IEEE HPSR 2021* by Hemant Rathore and Ashu Sharma
6. Malware Analysis @ *TENCON 2019, Kochi, India* by Hemant Rathore and Ashu Sharma
7. Malware Detection using Machine Learning and Deep Learning @ *BDA 2018, NIT Warangal, India* by Hemant Rathore and Mohit Sewak

11 Online Format

The speakers are comfortable in delivering the workshop in the online format. In fact, the speakers have taken many academic classes/workshops/tutorials in online mode only in the last year due to COVID-19 restrictions. Also, the workshop can be conducted on Google Meet, Zoom Meet, Cisco Webex, etc. In the past, the workshop was highly interactive, with a lot of open-ended discussions.

12 Speaker Brief Biography:

- **Mohit Sewak** is an Artificial Intelligence and Cybersecurity researcher with over 15 years of experience in designing innovative AI software and solutions. Mohit holds more than a dozen patents across the US, India, and worldwide for innovative AI solutions that empower many international products. Mohit is the author of multiple AI book titles on topics including technologies like Deep Reinforcement Learning and Convolutional Neural Networks. Mohit's research is focused on designing AI-based malware and other advanced threat detection and protection systems. Currently, Mohit is serving as a Principal Data Scientist for Security & Compliance Research at Microsoft R&D.
- **Hemant Rathore** is currently working as Assistant Professor at the Department of CS and IS at BITS Pilani, Goa Campus, India. Before joining academics, he was working in the area of computer security for three years at Symantec, India. His Ph.D. is on the topic of Adversarial Robustness and Explainability in Malware Detection Models. His research interests are in the area of Malware Analysis, Network Security, Machine Learning, and Operating Systems. He has guided several undergraduate and postgraduate students in their independent research projects and published many research papers in reputed journals/conferences.

References

- [1] Michael Ligh, Blake Hartstein, and Steven Adair. *Malware Analyst's Cookbook and DVD: Tools and Techniques for Fighting Malicious Code*. John Wiley & Sons Inc, 2010.
- [2] Digit Oktavianto and I Muhandianto. *Cuckoo Malware Analysis*. Packt Publishing Ltd, 2013.

- [3] Ori Or-Meir, Nir Nissim, Yuval Elovici, and Lior Rokach. Dynamic malware analysis in the modern era—a state of the art survey. *ACM Computing Surveys (CSUR)*, 52(5):1–48, 2019.
- [4] Junyang Qiu, Jun Zhang, Wei Luo, L Pan, S Nepal, and Y Xiang. A survey of android malware detection with deep neural models. *ACM Computing Surveys (CSUR)*, 53(6):1–36, 2020.
- [5] Michael Sikorski and Andrew Honig. *Practical Malware Analysis: The Hands-On Guide to Dissecting Malicious Software*. No Starch Press, 2012.
- [6] Carsten Willems, Thorsten Holz, and Felix Freiling. Toward automated dynamic malware analysis using cwsandbox. *IEEE Security & Privacy*, 5(2):32–39, 2007.
- [7] Yanfang Ye, Tao Li, Donald Adjeroh, and S Sitharama Iyengar. A survey on malware detection using data mining techniques. *ACM Computing Surveys (CSUR)*, 50(3):1–40, 2017.