

Malware Detection

Hivan J. Sabr



Outline



- Introduction
- Problem
- Malware Detection Process
- Malware Detection Approaches
- Malware Detection Datasets
- Malware Detection Evaluation
- Conclusion

Introduction



- In recent years, almost every member of the society has been using the Internet for daily life
 - Social interactions
 - Online banking
 - Health related transaction
 - Marketing
 -
- Any software which intentionally executes malicious payloads on victim machines (computers, smart phones, computer networks, etc.) is considered as malware

Introduction (Continue.)



- There many type of Malware such as
 - Virus
 - Worms
 - Trojan Horse
 - Rootkit
 -
- To protect legitimate users and companies from malware, malware need to be detected
- Malware detection is the process of determining whether a given program has malicious intent or not.

Problem in Malware Detection



- Problem of detecting the malware is NP-complete
 - Therefore it is impossible to design an algorithm which can detect all malware.

Problem in Malware Detection



- New generation of malware use techniques
 - such as encryption, oligomorphic, polymorphic, metamorphic, stealth, and packing methods to
 - make detection process more difficult
 - Easily bypass firewalls, antivirus software.
- Almost impossible to detect all malware with single detection approach because the
 - Computational complexity of malware is not clear
 - The problem is proved to be **NP-complete**.

Malware Detections Process



Malware Detections Process

Malware Analysis

Malware Future
Extraction

Classification

Static

Dynamic

Malware Detection Approaches



Approaches

Signature-based

Behavior-based

Heuristic-based

Model checking-based

Deep learning-based

Cloud-based

Mobile-based

IoT-based

Malware Datasets



- NSL-KDD dataset (2009)
- Drebin dataset (2014)
- Microsoft malware classification challenge dataset (2015)
- ClaMP (Classification of Malware with PE headers) dataset (2016)
- AAGM dataset (2017)
- EMBER dataset (2018)

Malware Detection Evaluation



- Evaluated in term of
 - their accuracy to detect malware.
 - Accuracy (Acc), Precision (Pr), Recall (Re), and F1 score are the four main classification metrics as follow:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Pr} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Re} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Conclusion



- No method could detect all new generation and sophisticated malware.
- The number, severity, sophistication of malware attacks, and cost of malware inflicts on the world economy have been increasing exponentially.
- Datamining and ML, new technologies such as deep learning, cloud, mobile devices, and IoT-based detection schemas have become popular.

References



- Aslan, Ö.A. and Samet, R., 2020. A comprehensive review on malware detection approaches. *IEEE Access*, 8, pp.6249-6271.
- Hemalatha, J., Roseline, S.A., Geetha, S., Kadry, S. and Damaševičius, R., 2021. An efficient densenet-based deep learning model for malware detection. *Entropy*, 23(3), p.344.