Malware classifications

Name of student

Name of professor

University

Course

Date

**Introduction**

Malware are a type of software programs that are developed and intentionally meant to bring down a computer service. Most of the malware applications are malicious in nature and cause bigger harm to the target host.

Examples of malwares include viruses, worms, ransom ware, adware, spyware and bots. Viruses for instance establish themselves on the target drives and replicate themselves consciously on the target machine, they may also intergrade with the normal running of certain computer applications Egele et al (2008). Applications that might be infected include Microsoft Office, video and music players. Unlike viruses, worms do not spread onto the other parts of the systems, but they exploit unknown vulnerabilities indie the applications that they target.

These exploits have been exploited in the past before a have made organizations to lose a lot of money. Ransom ware s on the other hand grab a portion of a resource within the target machine and then encrypt them with some high level algorithm, only to require the user /victim to reals a given amount of money in order to have their data set free. Adware come in a variety and advertise to the user unwanted or unsolicited ads to the user. Most of these ads may be so annoying, but usually the user has no way of controlling them. Spywares secretly spy on their victims by collecting location, messages, social media applications and calls, then send these information to the attacker or. Most attackers have used these information to gain undue surveillance advantage over their victims.

However, over the years, it has become increasingly difficult in trying to correctly classify an attack as either belong to any of the above discussed categories. Most malwares will be identified by the nature of the attack and how the actual program will work. Whereas most of the malware program have a uniform kind of behavior, it’s therefore been easier to identify and classify as either virus or a worm. Nonetheless, technologies continue to change over the years and this is why there is need to classify them. With the correct machine learning algorithm, it is possible to identify and classify these malwares appropriately.

**The classification process.**

**Background problem.**

To clearly understand why malware classification is necessary, we first of all need to understand why malwares have been here in the first place. Information and data are the backbone of any thriving business. An organization is only as valuable as the data that it keeps and interacts with. Whereas there have been rules that define how data should be consumed and used, these places have not been enough to stop data from landing in the wrong hands.

Banks and banking applications have been attacked severally attackers and in the process have ended up losing unaccounted sums of money. A simple object such as a flash disk/drive can be left unattended to and inserted in the computer networks, the exploit ends up consuming all the required permissions and resources required Ucci (2019).

End users have also lost worthwhile and sensitive information such as credit card details and banking usernames and passwords. These are just but a few samples and cases as to why data privacy and protection practices need to remain top notch priority. To accurately curb the menace, then security experts have to be able correctly classify each and every malware attack that comes in Peltier (2005). This is where machine learning and artificial intelligence comes in.

To get a correct classification outcome, the following process may be followed.

**Variable definition and data collection:** These involve the process of identifying the characteristics of the data to be evaluated. These may include:

1. Target industry:

Such industries may include, the finance industry, the government, the health care or even education industry. The industry choses for analysis are those that are more like to be attacked.

1. Target hosts devices

It’s easier to classify aware attacks based on the hosts that they originate from. These include, scream size, technology type and software applications of that particular device.

1. Device locations

Locations are also important in assessing the nature of attack, this will help determine, the nature of the target victims as afar as location is concerned.

1. Nature of targets
2. Nature of attack

Some attacks are meant to destroy information, others to deface or simply to steal information.

**Data cleaning:**

Once collected, these data should be put under A/B testing. During the A/B testing process, 80% of the data shall be used for the training process whereas the rest 20% shall be used for testing purposes. This ensures that the machine learns as much as possible from the provided dataset.

Algorithm application is a close approach to conducting data classification. Some of the known algorithm include KNN and Random forest. These algorithms have been specifically tailored to classify objects based on the given degree of shared characteristics with the rest of the shared class. KNN algorithm for instance is very effective in classifying objects based on the pre-defined attributes.

**References**

Egele, M., Scholte, T., Kirda, E., & Kruegel, C. (2008). A survey on automated dynamic malware-analysis techniques and tools. *ACM computing surveys (CSUR)*, *44*(2), 1-42.

Ucci, D., Aniello, L., & Baldoni, R. (2019). Survey of machine learning techniques for malware analysis. *Computers & Security*, *81*, 123-147.

Peltier, T. R. (2005). *Information security risk analysis*. CRC press.

Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.