**Ideation Phase**

**Brainstorm & Idea Prioritization Template**

|  |  |
| --- | --- |
| Date | 19 September 2022 |
| Team ID | 591488 |
| Project Name | Malware Detection and classification |
| Maximum Marks | 4 Marks |

**TEAM MEMBERS NAME:**

**ATHARV**

**SHAZ**

**SOHAN**

**PRATHEEK**

**Step-1: Team Gathering, Collaboration, and Select the Problem Statement**

Malware is malicious software designed to infect a system and achieve various malicious purposes. Malware can steal or encrypt data, capture login credentials, and take other actions to profit the attacker or harm the target.

Malware detection uses various tools and techniques to identify the presence of malicious software on a system. By proactively working to remediate malware infections on its systems, an organization can limit the cost and impact they have on the business.

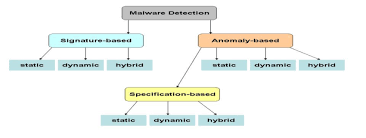
**Step-2: Brainstorm, Idea Listing and Grouping**

## Malware Detection Technologies

To implement these techniques and effectively detect malware, companies can use various tools, including:

* **Intrusion Detection System (IDS):**[An IDS](https://www.checkpoint.com/cyber-hub/network-security/what-is-an-intrusion-detection-system-ids/) is a security solution that identifies malware or other threats entering a network or installed on a system. An IDS generates an alert about the presence of the threat for security personnel to review.
* **Intrusion Prevention System (IPS):** [An IPS](https://www.checkpoint.com/cyber-hub/network-security/what-is-ips/)is similar to an IDS but takes a more proactive role in defending the organization against attack. In addition to generating an alert about identified threats, the IPS also blocks them from reaching the target system.
* **Sandboxing:** [Sandboxing](https://www.checkpoint.com/cyber-hub/threat-prevention/what-is-sandboxing/)involves performing dynamic analysis of malware in a safe, isolated environment. Malware sandboxes have various built-in tools designed to monitor the malware’s activities, determine if it is malicious, and map out its capabilities.
* **Malware Analysis Tools:** Malware analysis tools are available to implement the various malware detection techniques described previously. For example, disassemblers like the Interactive Disassembler (IDA) are used for static analysis, while a debugger is a common tool for dynamic analysis.
* **Cloud-Based Solutions:** Cloud-based infrastructure provides organizations with the ability to enhance their malware detection capabilities beyond what is feasible in-house. Cloud-based solutions can distribute IoCs to the users of a particular solution and perform sandboxed analysis of potential malware at scale.

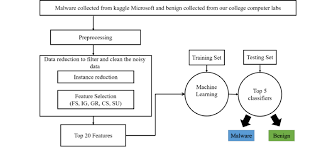
**Step-3: Idea Prioritization**



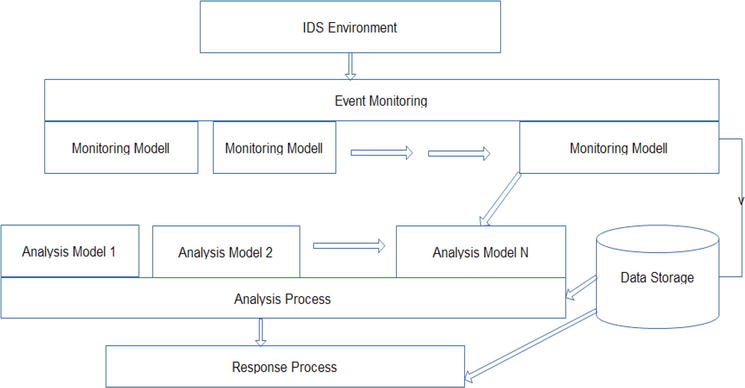
illustrates the techniques for malware detection that work in a flow including data processing, feature selection, classifier training, and malware detection. The process begins with collecting datasets from the Kaggle website consisting of malware and benign web applications. By adopting AI technology, the development of malware detection systems shall be in a way that will process malware datasets, and analyze malware to understand its feature. Fisher Score (FS), Chi-Square (CS), Information Gain (IG), Gain Ratio (GR), and Uncertainty Symmetric (US) are used to select 20 features. The system shall train the classifier by comparing different classifiers on FS, CS, IG, GR, and US to detect unknown malware.

Implementing different types of classifiers to develop malware detection and prevention systems shall provide better and using AI shall bring a significant advantage to detect and prevent unknown malicious activities [35]. In Fig. 5, we display a flowchart of unknown malware detection using artificial intelligence. In this section, we provide a detailed review of each method of Malware Detection.

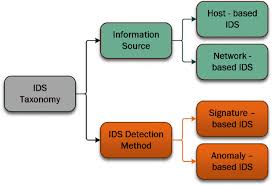
• Signature-based Detection Technique: The signature-based detection method consists of four components as depicted in Fig. 6 is a term that helps in identifying and detecting attacks by looking for specific patterns [40]. In a signature-based method, developers use a database containing signatures of viruses, scan the file, and evaluate information with that database for detecting malware in the database. If the information matches with the database’s data that means the file contains viruses. The primary advantage of this method is effective for the known malware, however, it has limitations in detecting unknown malware [41]. Fig. 7 shows Intrusion Detection System (IDS) keeps a statistical model of traffic that also can be referred to a database, IDS accepts traffic from various sources and matches it with statistical traffic to find out whether it is malicious or not and then provides the result to an administrator.



* Anomaly-based Detection Technique: Anomaly-based network intrusion detection plays a vital role in addressing security issues and protecting networks against malicious activities [43]. Anomaly-based methods address the limitations of signature-based techniques by enabling to detect any of known or unknown malware by applying classification techniques over activities of a system for malware detection. Such transformation from pattern based detection to a classification-based approach to identify normal or anomalous behavior gives an advantage of detecting malware activities



It depicts the anomaly-based Network Intrusion Detection System (IDS) where the functional stages are normally adopted in the anomaly-based network intrusion detection systems (ANIDS)



On the other hand, Fig.illustrates a connection with a database consists of the signature of known attacks, with the common signatures coming from different packets with that database, an alert is sent to the system admin if the unknown signature matches with known signature mean malware detected.

* Heuristic-based Detection Technique: Applying Artificial Intelligence over the signature and anomaly-based detection systems improve the efficiency of malware detection. However, in order to adopt environmental change and improve prediction ability, a machine learning algorithm named genetic algorithm along with neural network was applied over malware detection system to improve the classification method. The algorithm applies characteristics such as inheritance, selection, and combination that give the advantage to attain optimum solutions from multiple directions without any previous knowledge about the system [45]. The combination of statistical and mathematical techniques improves the heuristic method from previous methods. It represents the features of the Heuristic Method

