Project Design Phase-I
Solution Architecture

Date	30 October 2023
Team ID	Team 2.11
Project Name	Malware Detection and Classification
Maximum Marks	4 marks

Solution Architecture:

Tech Solution:

- Data Collection: The first step in the malware detection and classification process is to collect data. We are using datasets from Kaggle. It involves converting the data into a format that can be used by your machine learning models.

- Feature extraction: The next step is to extract features from the data. Features are characteristics of the data that can be used to distinguish between malicious and benign samples. Some common features for malware detection include file size, file type, number of imports, number of exports, and presence of known malware strings.

- Model training: Once the features have been extracted, we can train our machine learning models. There are a variety of machine learning algorithms that can be used for malware detection and classification, such as support vector machines (SVMs), decision trees, and random forests.

- Model evaluation: Once the models have been trained, we need to evaluate their performance on a held-out test set. This will help us to identify the best model and to tune the hyperparameters of the model.

- Model deployment: Once we have identified the best model, we can deploy it to production. This may involve integrating the model into a security information and event management (SIEM) system or other security solution.

Software structure:

A typical software architecture might include the following components:

- Data collection module: This module is responsible for collecting malware samples and other relevant data.
- Feature extraction module: This module extracts features from the collected data that can be used to detect and classify malware.
- Detection and classification module: This module uses machine learning algorithms to detect and classify malware samples based on the extracted features.
- Reporting module: This module generates reports on the detected malware samples, including their classification and other relevant information.

Software characteristics:

The following are some important characteristics of malware detection and classification software:

- Accuracy: The software should be able to accurately detect and classify malware samples, even in the presence of new and unknown malware.
- Speed: The software should be able to detect and classify malware samples quickly, so that you can take action to mitigate the risk posed by the malware.
- Scalability: The software should be scalable to handle large volumes of malware samples and data.
- Ease of use: The software should be easy to use and manage, even for users with limited technical expertise.

Stakeholder communication:

It is important to communicate effectively with our project stakeholders to ensure that they understand the structure, characteristics, behaviour, and other aspects of your malware detection and classification software. Some tips for stakeholder communication include:

- Use clear and concise language.
- Avoid technical jargon.
- Use diagrams and illustrations to explain complex concepts.
- Be prepared to answer questions.

Feature definition:

The following are some important features we are considering for your malware detection and classification software:

- Ability to detect a wide range of malware: The software should be able to detect a wide range of malware, including viruses, worms, Trojans, spyware, and ransomware.
- Ability to detect new and unknown malware: The software should be able to detect new and unknown malware, even if the malware has not been previously seen.
- Ability to classify malware accurately: The software should be able to classify malware accurately, so that you can take appropriate action to mitigate the risk posed by the malware.
- Ability to generate detailed reports: The software should be able to generate detailed reports on the detected malware samples, including their classification and other relevant information.

Development phases:

The following are some typical development phases for malware detection and classification software:

- Requirements gathering: This phase involves gathering requirements from the project stakeholders and defining the features and functionality of the software.
- Design: This phase involves designing the architecture of the software and developing a plan for implementing the features.
- Implementation: This phase involves implementing the software according to the design plan.
- Testing: This phase involves testing the software to ensure that it meets the requirements and works as expected.
- Deployment: This phase involves deploying the software to production environment and making it available to users.

Solution requirements:

- Performance requirements: The system should be able to detect and classify malware samples quickly, so that we can take action to mitigate the risk posed by the malware.
- Accuracy requirements: The system should be able to detect and classify malware samples with a high degree of accuracy.
- False positive rate requirements: The system should have a low false positive rate, so that we are not unnecessarily alerted to non-malicious files.
- False negative rate requirements: The system should have a low false negative rate, so that we do not miss any malicious files.
- Cost requirements: The system should be cost-effective to implement and maintain.

The following are some specifications for your malware detection and classification project, which you can use to define, manage, and deliver the solution:

Functional specifications:

The functional specifications should describe the features and functionality of the system, including the following:

- What types of malware can the system detect and classify?
- How does the system collect data on malware samples?
- What features are extracted from the data?
- What machine learning algorithms are used to detect and classify malware?
- How does the system generate reports on detected malware samples?

Non-functional specifications:

The non-functional specifications should describe the performance, scalability, reliability, security, and usability requirements of the system, including the following:

- What are the performance requirements for the system? (e.g., how quickly must the system be able to detect and classify malware samples?)
- What are the scalability requirements for the system? (e.g., how many malware samples per day must the system be able to handle?)
- What are the reliability requirements for the system? (e.g., what is the acceptable downtime for the system?)
- What are the security requirements for the system? (e.g., how will the system protect the confidentiality and integrity of the data it processes?)
- What are the usability requirements for the system? (e.g., how easy must the system be to use and manage for users with limited technical expertise?)

Management specifications:

The management specifications should describe the processes and procedures for managing the development and delivery of the solution, including the following:

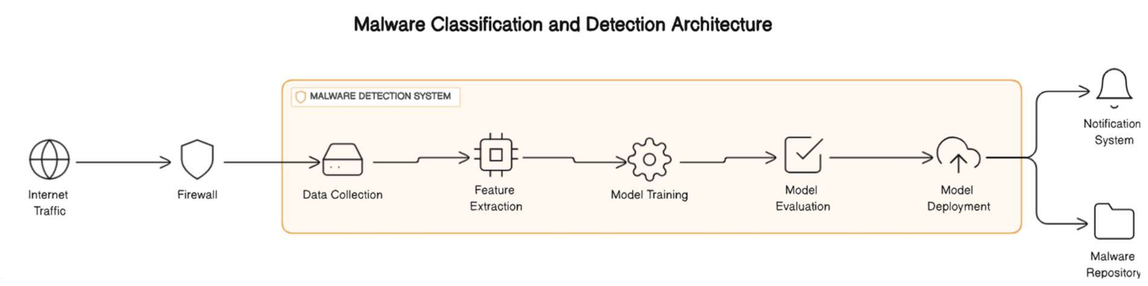
- What are the development phases for the solution?
- What are the roles and responsibilities of the different stakeholders involved in the project?
- What are the communication and change management procedures for the project?
- How will the risks for the project be managed?

Delivery specifications:

The delivery specifications should describe the processes and procedures for delivering the solution to the customer, including the following:

- What are the acceptance criteria for the solution?
- How will the solution be deployed and tested?
- How will the customer be trained on how to use the solution?
- How will the solution be supported after deployment?

Example - Solution Architecture Diagram:


Malware Classification and Detection Architecture
MALWARE DETECTION SYSTEM
Internet Traffic
Firewall
Data Collection
Feature Extraction
Model Training
Model Evaluation
Model Deployment
Notification System
Malware Repository