Ministry of Higher Education and

Scientific Research

University of Babylon

College of Information Technology

Department of Information Security

Study: (Morning, Evening)

Extracting and Clustering Malware With K-Means

A Graduate Project Submitted to the department of Information Security of the College of Information Technology, University of Babylon, in Partial Fulfillment of the Requirements for the Bachelor's degree in the Information Security of Information Technology.

By

Ahmed Abd Al-Kareem Ali

Supervised by

Assist. Prof.Dr.Ahmed Khelfa Al-Ajeli

2023-2024

Abstract

The utilization of machine learning techniques for malware detection has become increasingly crucial in contemporary cybersecurity landscapes. In this project, we delve into the realm of malware classification by employing a KMeans clustering model on features extracted from Portable Executable (PE) files. These features encompass a spectrum of attributes such as file size, header information, and byte-level characteristics, which collectively form a comprehensive representation of each file's intrinsic properties.

Despite the overarching goal of accurately distinguishing between malicious and Normal PE files, our implementation confronts several challenges and opportunities for refinement. One notable issue lies in the selection and extraction of features, as the efficacy of the clustering model heavily relies on the discriminative power of these attributes. Ensuring the inclusion of relevant features while mitigating noise and redundancy constitutes a pivotal aspect of our endeavor.