République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieure et de la Recherche Scientifique
Ecole Supérieure en Informatique de Sidi Bel Abbès (ESI-SBA)


**Mémoire de Fin d'étude**

Pour l'obtention du diplôme d'**ingénieur d'état**

Filiére: **Informatique**

Spécialité: **Ingénierie des Systèmes Informatiques (ISI)**

# Thème

# Detection of malicious PowerShell scripts using machine learning and deep learning

Réalisé par:

- MERZOUK BENSELLOUA Ahmed Yasser
- MESSADI Said Abdesslem

Année Universitaire

2022/2023

# Abstract

This thesis addresses the challenge of detecting malicious PowerShell scripts using machine learning and deep learning techniques. We conduct a comprehensive review of the state of the art and identify the limitations of existing methods. Our research focuses on the application of Large Language Models (LLMs), such as BERT, which demonstrate remarkable capabilities in capturing contextual information and semantic dependencies. We experiment with various models, including Bidirectional LSTM (BLSTM), and develop a comprehensive solution that includes an event log consumer, a high-performance API, and a user-friendly web application. Through extensive evaluation, we achieve highly accurate detection results, highlighting the potential of machine learning and deep learning in combating PowerShell-based cyber threats. This thesis contributes valuable insights and practical techniques for researchers and practitioners in the field.

# Introduction

PowerShell, a powerful scripting language and automation framework developed by Microsoft, has become a double-edged sword in the realm of cybersecurity. While PowerShell offers immense flexibility and productivity for system administrators and IT professionals, it has also emerged as a preferred tool for malicious actors to carry out cyber attacks. Detecting and mitigating the risks associated with malicious PowerShell scripts have become critical tasks in the modern cybersecurity landscape.

In this thesis, we delve into the realm of PowerShell script detection and explore novel techniques that leverage machine learning and deep learning algorithms. Our objective is to develop robust and effective methods to distinguish between benign and malicious PowerShell scripts, enabling organizations to proactively identify and thwart potential cyber threats. We begin by conducting a comprehensive review of the state of the art, examining existing approaches and their limitations.

Drawing inspiration from the advancements in Natural Language Processing (NLP) and the success of Large Language Models (LLMs), such as BERT, we explore their applicability in the context of PowerShell script analysis. These LLMs have demonstrated exceptional capabilities in capturing contextual information and semantic dependencies, making them promising candidates for detecting malicious PowerShell code. Furthermore, we investigate the effectiveness of traditional machine learning algorithms, such as Bidirectional LSTM (BLSTM), which have shown considerable potential in sequence classification tasks.

To validate and evaluate our proposed techniques, we construct a comprehensive experimental setup that includes data collection, preprocessing, model training, and performance evaluation. Additionally, we develop a robust event log consumer that interfaces with our models through a high-performance API built using the FastAPI

framework. The output of our detection models is seamlessly integrated into a user-friendly web application, providing detailed insights into script classifications and historical detection results.

Through rigorous experimentation and evaluation, we demonstrate the efficacy of our proposed methods in accurately detecting malicious PowerShell scripts. Our research contributes valuable insights and practical techniques that can aid cybersecurity researchers and practitioners in developing proactive defense mechanisms against PowerShell-based cyber threats. By addressing the challenges posed by malicious PowerShell scripts, we aim to enhance the security posture of organizations and protect critical systems from potential breaches and data exfiltration.