Team Name: MalWhere

Project Title: AI-powered Cybersecurity System

---

1. Introduction

With the increasing number of cyber threats, detecting anomalies, malware, and phishing attempts is crucial to securing digital environments. Our project leverages Azure AI services to develop a robust cybersecurity system capable of detecting network intrusions, malware, and phishing attempts in real time.

---

2. Problem Statement

Cyber threats such as network intrusions, malware infections, and phishing attacks pose a significant risk to organizations. Existing solutions often struggle with scalability, real-time detection, and false positives. Our AI-powered system addresses these challenges using advanced anomaly detection, text analytics, and computer vision techniques.

---

3. Objectives

- **Anomaly Detection (NIDS):** Identifies abnormal activities in network traffic logs.
- **Malware Detection:** Detects malicious files and programs.
- **Phishing Detection:** Identify phishing emails and malicious URLs.
- **Cybersecurity Awareness Chatbot:** Educates users about cybersecurity threats, best practices, and real-time threat detection guidance.

---

4. Solution Overview

Our AI-powered cybersecurity system utilizes Azure AI services to analyze and classify network traffic, files, and emails in real-time. The system consists of four main components:

4.1 Anomaly Detection (NIDS)

- **Packet Capture:** The system captures live network packets using Scapy, continuously monitoring both incoming and outgoing traffic.
- **Feature Extraction:** It extracts key packet features such as source and destination IP, port, protocol, and size for further analysis.
- **Machine Learning Classification:** A **RandomForestClassifier**, trained on the **USTCTFC2016 dataset**, is used to classify packets as **benign** or **malicious** based on extracted features.
- **Threat Detection:** The system detects **malware-infected traffic in real time**, enhancing network security by identifying suspicious behavior early.

4.2 Malware Detection

- **Feature Extraction:** Extracts key features from **Portable Executable (PE) files** using the **pefile** library.
- **ML-Based Classification:** Sends extracted features to a **deployed API**, which uses a **Random Forest Classifier** for malware detection.
- **Backend Processing:** Implements a **FastAPI-based backend** to handle predictions and return classification results.
- **Cloud-Hosted Model:** Utilizes a **trained Random Forest model stored in Google Drive** for classifying files as **malware or safe**.
- **Automated Scanning:** Runs a **monitoring service** that continuously scans new files in the user's **Downloads folder** for potential threats.

4.3 Phishing Detection

- **Automated Email Scanning:** Continuously monitors **incoming emails** for potential phishing threats without user intervention.
- **Smart Email Categorization:** Differentiates between **important, safe, and unsafe emails** to enhance inbox security.
- **Threat Logging & Analysis:** Maintains **detailed logs of flagged emails**, allowing security teams to review potential phishing attempts.

- **User Security Recommendations:** Provides actionable steps to help users **safeguard themselves from phishing threats**.

4.4 ChatBot

- **AI-Powered Responses:** Utilizes Azure OpenAI's GPT-4 for advanced natural language understanding and contextual cybersecurity guidance.
- **Incident Assistance:** Helps users handle security incidents by providing step-by-step solutions and best practices.
- **Security Best Practices:** Advises users on password security, multi-factor authentication (MFA), and safe browsing habits.
- **24/7 Cybersecurity Support:** Provides round-the-clock assistance for cybersecurity-related queries and concerns.
- **Vulnerability Awareness:** Educates users on common cyber threats and offers guidance on securing systems against exploits.

---

5. Technologies and Tools

- **Azure AI Services:** Azure OpenAI (GPT-4, Codex, DALL-E), Azure Anomaly Detector, Azure Cognitive Services (Text Analytics, Computer Vision)
- **Visual Studio Code and Render:** For training and deploying detection models.
- **Machine Learning Stack:** Python, Jupyter Notebook, TensorFlow/PyTorch, Scikit-learn.
- **PE File Analysis:** pefile for extracting features.
- **Scapy:** captures live network packets from a specified interface

---

6. Deployment Guide

1. Set up Azure Storage and AI Services.
2. Deploy AI models using Azure Machine Learning.
3. Add a credentials.json file in phishingDetection folder.
4. Add this env file in fishingDetection folder
    a. `AZURE_OPENAI_ENDPOINT`

    b. AZURE_API_KEY
    c. FETCH_EMAILS_URL
    d. AZURE_OPENAI_DEPLOYMENT
5. Add this env file in ChatBot folder
    a. AZURE_OPENAI_ENDPOINT
    b. AZURE_API_KEY
    c. AZURE_OPENAI_DEPLOYMENT
6. **Run the FastAPI server for malware classification:**
   uvicorn server:app --host 0.0.0.0 --port 8000
7. **Run the file monitoring script:**
   python MalwareDetectionInFiles/detectMalware.py
8. **Run the network packet monitoring script:**
   python networkIntrusion/demo.py
9. **Run the Email Fetching file for access to fetching user email:**
   python phishingDetection/emailFetching.py
10. **Run the Phishing Detection file:**
    python phishingDetection/server.py
11. **Run the ChatBot Server file:**
    chatbot server.py-> uvicorn server:app --host 127.0.0.1 --port 8000
12. Deploy the application on Azure App Service or a cloud-based platform.
13. Monitor and refine model performance over time.

8. Testing and Evaluation

- **Dataset Used:** Public cybersecurity datasets.
- **Metrics Evaluated:**
    - **Accuracy & Precision:** Measures the correctness of threat detection.
    - **Recall:** Ensure the threat doesn't go undetected.
    - **Inference Time:** Ensure real-time detection capabilities.
- **Testing Results:** Evaluation performed using standard cybersecurity datasets.

9. Challenges Faced

- Handling imbalanced datasets for anomaly detection.
- Reducing false positives in phishing detection.
- Ensuring real-time processing without performance bottlenecks.

10. Ethical Considerations

- **Bias Mitigation:** Ensuring fair and unbiased AI decision-making.
- **Data Privacy & Compliance:** Following GDPR and Microsoft Responsible AI guidelines.
- **Explainability:** Providing clear reasons for threat detection to end-users.

11. Future Improvements

- Expanding dataset coverage for improved accuracy.
- Enhancing phishing detection with advanced deep learning techniques.
- Implementing blockchain-based threat intelligence sharing.

12. Conclusion

Our AI-powered cybersecurity system is innovative in detecting network intrusions, malware, and phishing threats. By integrating Azure AI services and machine learning-based malware detection, our solution enables organisations to enhance their security posture with real-time, AI-driven threat detection.