Digital Detectives

Owner: Team 1
Reviewer: Team 1
Contributors:
Date Generated: Sun Dec 08 2024

Executive Summary

High level system description

This application is a local, LLM-based digital forensic tool designed to assist users in analyzing disk images for potential threats and generating insights. The core functionality involves the following steps:

User Interaction: A user uploads a disk image via a Web UI.
Forensic Analysis: The uploaded disk image is processed locally using The Sleuth Kit, and file analysis results are sent to VirusTotal for malware detection.
Data Management: The analysis generates a dataset stored in both a CSV file and an SQL database for further queries.
LLM Integration: Users can query the dataset through the Web UI, and the application employs a Local LLM to answer user questions based on the analysis results.

Core Components and Interactions

External Elements:
User: Interacts with the application through the Web UI.
VirusTotal: Provides external malware detection services.

Internal Elements:
Web UI: The primary interface for user interaction.
LLM: Processes user queries and generates answers from the dataset.
Analysis Module: Performs forensic analysis on the uploaded disk image and connects with VirusTotal.
Dataset: Stores analysis results in a CSV file and SQL database, accessible by the Analysis Module and LLM.

Trust Boundaries

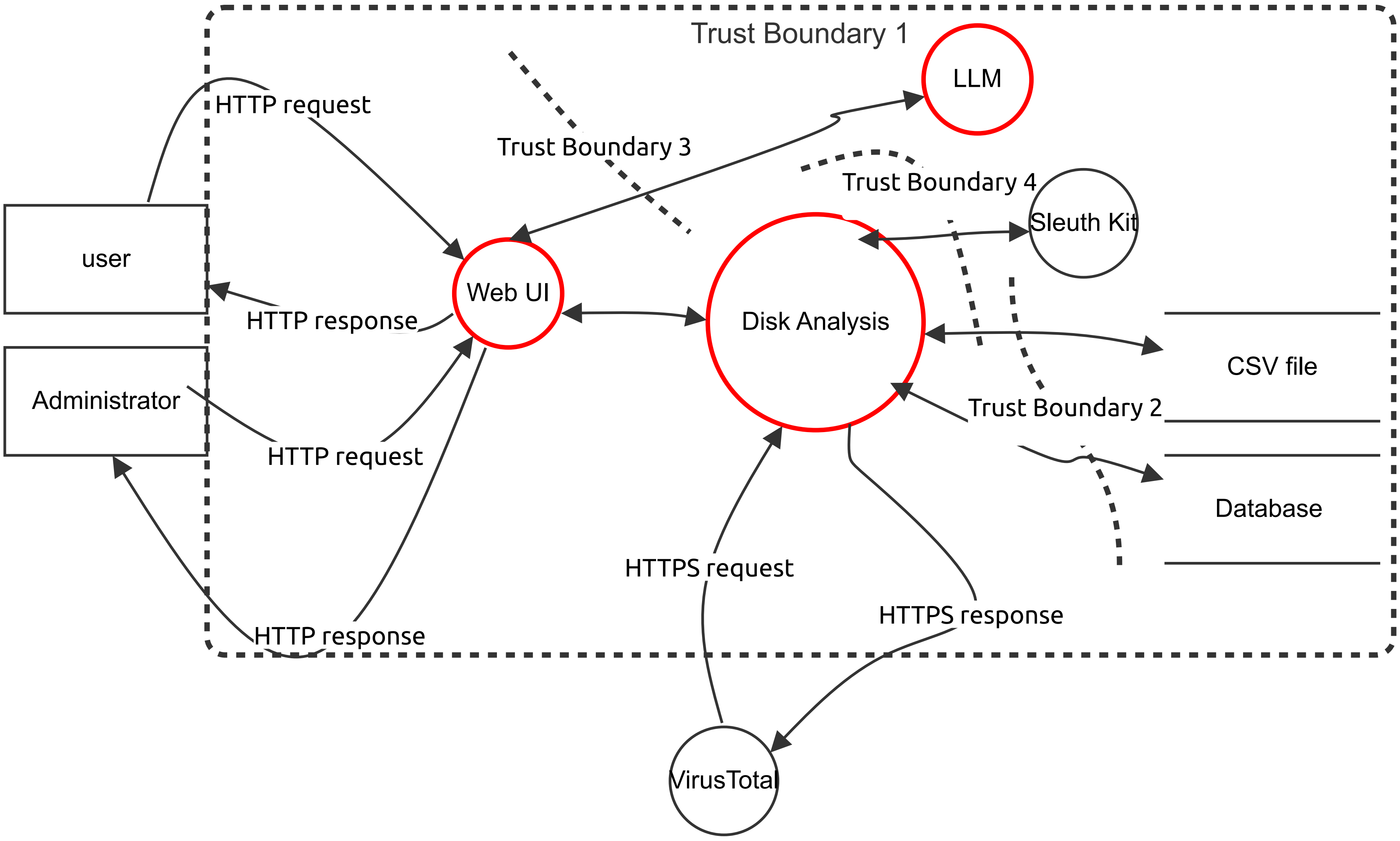
Trust Boundary 1: Separates external elements (User, VirusTotal) from internal components (Web UI, Analysis Module, Dataset, LLM).
Trust Boundary 2: Exists between the Web UI and the Analysis Module to delineate analysis functionality from the user interface.
Trust Boundary 3: Exists between the Web UI and the LLM to ensure controlled interaction between user queries and the LLM.

Summary

Total Threats	15
Total Mitigated	0
Not Mitigated	15
Open / High Priority	0
Open / Medium Priority	5
Open / Low Priority	0
Open / Unknown Priority	0

Digital Detectives

LLM powered web application



Digital Detectives

user (Actor)

low privileged, Interacts with the application through the Web UI.

Number	Title	Type	Priority	Status	Score	Description	Mitigations
26	Spoofing	Spoofing	Medium	NotApplicable		Changing a small part of, for example malware, the virustotal check could also not yield a legitimate result and pass it as non detected(Obfuscation)	Running the files(not only sending hashes to VT)
27	Repudiation	Repudiation	Medium	NotApplicable		Provide a description for this threatA malicious user may change or forge data from the program to fake the content of the disk image.	User based access control with MFA for all users. Logging of user activity.

Administrator (Actor)

High privileged , Interacts with the application through the Web UI.

Number	Title	Type	Priority	Status	Score	Description	Mitigations
28	Spoofing	Spoofing	Medium	NotApplicable		Changing a small part of, for example malware, the virustotal check could also not yield a legitimate result and pass it as non detected(Obfuscation)	Running the files(not only sending hashes to VT)
29	Repudiation	Repudiation	Medium	NotApplicable		A malicious user may change or forge data from the program to fake the content of the disk image.	User based access control with MFA for all users. Logging of user activity.

Web UI (Process)

The primary interface for user interaction.

Number	Title	Type	Priority	Status	Score	Description	Mitigations
22	Spoofing	Spoofing	Medium	Open		Changing a small part of, for example malware, the virustotal check could also not yield a legitimate result and pass it as non detected(Obfuscation)	Provide remediation for this threat or a reason if status is N/A
23	Denial of service	Denial of service	Low	NotApplicable		If the application was deployed on a network more people could possibly try to use it which it is not designed for. To hide evidence and make it hard for the program to handle the diskimage, use of large file names and paths could be used	The application is used locally and one user at a time using the application
24	Information disclosure	Information disclosure	Medium	NotApplicable		A user could possibly access other reports and disk images which it should not be allowed to.	Not applicable because the application is used locally.
25	New STRIDE threat	Elevation of privilege	Medium	NotApplicable		Only username and password. (brute force attacks). A user accessing reports which it should not be allowed to.(program is run locally)	ACL, ASLR, running application with the least privilege possible.

Disk Analysis (Process)

Performs forensic analysis on the uploaded disk image and connects with VirusTotal.

Number	Title	Type	Priority	Status	Score	Description	Mitigations
30	Tempering	Tampering	Medium	Open		Changing or removing the disk images would have a severe impact on the integrity of the diskimages. As this is an investigation tool this would render the diskimage useless.	Keeping backups of disk images and reports not accessible to normal users and at several locations. Hashing all reports and disk images for accounting. Sending a copy(which the user cannot access) of reports to another location. Logging user activity.

CSV file (Store)

Stores analysis results in a CSV file and SQL database, accessible by the Analysis Module and LLM

Number	Title	Type	Priority	Status	Score	Description	Mitigations

Database (Store)

Stores analysis results in a CSV file and SQL database, accessible by the Analysis Module and LLM

Number	Title	Type	Priority	Status	Score	Description	Mitigations

HTTP request (Data Flow)

Low privileged

Number	Title	Type	Priority	Status	Score	Description	Mitigations

HTTP request (Data Flow)

High privileged

Number	Title	Type	Priority	Status	Score	Description	Mitigations

HTTP response (Data Flow)

High privilged

Number	Title	Type	Priority	Status	Score	Description	Mitigations

HTTP response (Data Flow)

High privileged

Number	Title	Type	Priority	Status	Score	Description	Mitigations

Data Flow (Data Flow)

Number	Title	Type	Priority	Status	Score	Description	Mitigations

Data Flow (Data Flow)

Write and read to and from CSV file

Number	Title	Type	Priority	Status	Score	Description	Mitigations

Data Flow (Data Flow)

Write and read from and to Database

Number	Title	Type	Priority	Status	Score	Description	Mitigations

Data Flow (Data Flow)

Number	Title	Type	Priority	Status	Score	Description	Mitigations

Data Flow (Data Flow)

Number	Title	Type	Priority	Status	Score	Description	Mitigations

HTTPS response (Data Flow)

Number	Title	Type	Priority	Status	Score	Description	Mitigations

HTTPS request (Data Flow)

Number	Title	Type	Priority	Status	Score	Description	Mitigations

Sleuth Kit (Process)

Number	Title	Type	Priority	Status	Score	Description	Mitigations

LLM (Process)

Processes user queries and generates answers from the dataset.

Number	Title	Type	Priority	Status	Score	Description	Mitigations
11	Training Data Poisoning	Tampering	Low	NotApplicable		Training data poisoning involves altering the data used to train the model, which constitutes unauthorized modification of assets (data tampering).	Not relevant, usage of pre-trained model and not performing further training.
12	Prompt Injection	Tampering	Medium	Open		Prompt injection exploits LLMs by manipulating prompt to produce unintended response	Input Filtering; Define a whitelist of acceptable terms, keywords or query structures allowed in user prompt. For example only queries related to file attributes, metadata, or search terms presents in the CSV should be allowed. Contextual Validation; validate inputs to ensures they're within forensic context.
13	Jailbreaking	Elevation of privilege	Medium	NotApplicable		y "jailbreaking" the LLM, attackers might bypass security mechanisms and gain unintended access to restricted functionalities or capabilities.	Provide remediation for this threat or a reason if status is N/A
14	DOM-Based Attacks	Tampering	Low	NotApplicable		These attacks involve embedding harmful instructions within HTML or scripts, altering the intended behavior of an application that processes this content.	Not relevant, because not scraping external website
15	Denial of Service	Denial of service	Medium	Open		This directly corresponds to the STRIDE category for attacks aimed at disrupting availability and causing performance degradation.	Not apply to our case, because it is a local based service, and only one user using it at a time
16	Data Leakage	Information disclosure	Medium	Open		Data leakage involves the inadvertent exposure of sensitive information, which aligns with the STRIDE category for breaches in confidentiality.	Not applicable because it application is used for forensic purpose, and the analyst needs all the information when analysing data for evidence.

VirusTotal (Process)

Provides external malware detection services.

Number	Title	Type	Priority	Status	Score	Description	Mitigations