Team404

Prevent, Mitigate, and Recover (PMR) Insight Collective Knowledge System (PICK) Tool Software Design Document

Version 1.0

March 09, 2020

Document Control

Approval

The Guidance Team and the customer shall approve this document.

Document Change Control

Initial Release:	0.1
Current Release:	1.0
Indicator of Last Page in Document:	**$**
Date of Last Review:	07 March 2020
Date of Next Review:	11 March 2020
Target Date for Next Update:	12 March 2020

Distribution List

This following list of people shall receive a copy of this document every time a new version of this document becomes available:
Guidance Team Members:
Dr. Ann Gates
Dr. Salamah Salamah
Dr. Steven Roach
Elsa Tai Ramirez
Peter Hanson
Jake Lasley

Customer:
Dr. Oscar Perez
Vincent Fonseca
Herandy Denisse Vazquez
Baltazar Santaella
Florencia Larsen
Erick De Nava

Software Team Members:
Eduardo A Jiménez Todd
Jacob N Torres
Jorge I Felix
Alejandro Zamora
Matthew S Montoya

Change Summary

The following table details changes made between versions of this document

Version	Date	Modifier	Description
0.1	26 February 2020	Matt Montoya	Creation of Document
0.2	06 March 2020	Alex Zamora	Uploaded Database Schema based on ER diagrams by Jorge & Matt

0.3	07 March 2020	Jorge Felix	Finalized Collaboration Diagram with the help of Eddy & Jacob
0.4	08 March 2020	Eddy Todd	Finished Introduction with the help of Jacob & Matt
0.5	09 March 2020	Jacob Torres	Fixed Collaboration Diagram
1.0	09 March 2020	Matt Montoya	Fixed grammar errors, updated DB schema

Table of Contents

Introduction 1
Purpose and Intended Audience 1
Scope of Product 1
References 1
Definitions, Acronyms, and Abbreviations 1
Definitions 1
Acronyms 3
Abbreviations 3
Overview 3
Decomposition Description 4
Detailed Description of Components 4
Database 4

Decomposition Description 5
SYSTEM COLLABORATION DIAGRAM 5
Subsystem and Component Descriptions **Error! Bookmark not defined.**
Dependencies 12

Detailed Description of Component Splunk 13
Component Description **Error! Bookmark not defined.**
CLASS DESCRIPTION SPLUNK 13
Contract Provide Log Files 13

DATABASE 14
Database Schema 14

1. Introduction

Section 1 shall introduce Team404's Software Design Document (SDD) document for the Spring 2020 Software II Project, *PICK Tool*, including the purpose and intended audience, overview, document references, as well as definitions, acronyms, and abbreviations.

1.1. Purpose and Intended Audience

The purpose of creating the software design document is to describe PICK tool, in order to give the software development team general guidance for the architecture of the project. This Software Design Document details the specifications of the characteristics of the design components. The description of the design is necessary to coordinate the team under a single vision. It must act as a point of reference and outline all parts of the software and how they work between them.

1.2. Scope of Product

PICK Tool shall facilitate the job of analysts during an Adversarial Assessment of a simulated cyber-attack, reducing the time it currently takes analysts to perform an assessment to about two weeks. In doing so, PICK Tool shall assist analysts in telling the true story pertaining to these simulated attacks. To help satisfy these needs, analysts will utilize PICK Tool to search through and filter through logs, or recorded notes from the systems of attackers and defenders, as they may pertain to a simulated attack. Though PICK Tool, analysts can use PICK Tool to construct a vector or visual graph of events that satisfy an objective. More information regarding logs can be found in Section 2 of this document.

1.3. References

[1] Wirfs-Brock, R., Wilkerson, B. and Wiener, L. (1990). Designing object-oriented software. Englewood Cliffs, N.J.: Prentice-Hall.

1.4. Definitions, Acronyms, and Abbreviations

1.4.1. Definitions

TERM	DEFINITION
Active Scene	The scene that is currently displayed.
Actor	A user or external system that interacts with the system in the use case diagram.
Adversarial Assessment	Analysis of a simulated cyberattack by the White Team.
Analyst	The Analyst is the primary user of PICK Tool. Multiple Analysts can access PICK Tool simultaneously. The analyst primarily uses PICK Tool to ingest log files, correlate logs, and create graphs within the system.
Associated Log	Logs with a cause and consequence log linked together.
Audio Transcription Tool	A software program that takes audio logs and transcribes them into text files.
Blue Team	Defenders during the simulated cyber-attack; the team that will defend their system from attack.
Client(s)/Customer(s)	The U.S. Army Combat Capabilities Development Command Data & Analysis Center: Lethality, Survivability & Human System Integration (LSH) Directorate; individuals from this directorate include Dr. Oscar Perez, Mr. Vincent Fonseca, Mr. Baltazar Santaella, Ms. Herandy Vazquez, Ms. Florencia Larsen, & Mr. Erick De Nava.

Client-Server Model	PICK Tool running on a server in a closed system.
Commit	A button that confirms the actions of the analyst to save the project.
Correlated Logs	Two log entries connected through cause and effect by the analyst.
Database	A structured set of data held in a computer, especially one that is accessible in various ways.
ETL Tool/Log Management Tool	Software that combines the Extract, Transform, Load database functions into one tool to take data from one or many sources into a destination system.
Event Log	A detailed record of system events stored in the System Event Viewer by the computer's operating system.
Filter	A way for the analyst to search for specific queries through specific conditions set by the analyst.
Filter Space	A certain condition set by the analyst.
Formatted Log Entry	A sanitized log entry that has been cleaned up based on a set configuration by the analyst.
Graph	A visual representation of the scenario between the BlueTeam and the Red Team shown through nodes and vectors. [1]
Graphing Tool	A software program that will construct the graph based on the vectors and correlations made by the analyst.
Graphical User Interface	The way that the system will display all components of the system to the user, the analyst.
Ingest/Ingestion	When files, specifically formatted log files, are put into our system to be correlated, edited and disregarded if need be.
Kali Linux	Operating System the clients will use.
Local Network	A data communications network within the system.
Log	An official record of events.
Log Entry	A log entry is the output of the log file. Log entries contain information from the log file and are recorded details of an ingested log.
Log File	A file is an input into the system that records either events that occur in an operating system or other software runs or messages between different users of communication software.
Natural Language Processor	An external software program that takes audio files and translates them into text files to be later validated.
Node	A node is a visual representation of a significant event that was marked for the current vector.
Normalize	To make rid of duplicate entries & make equal amongst all other entries.
Objective	The goal of the attackers when they attack a system. It is the goal of the defenders to prevent attackers from achieving their goal(s).
Observer	A member of the white team that takes notes and assessments during the staged cyber-attack between the Blue Team and the Red Team.
Optical Character Reader	An external software program that takes image files and translates them into text files to later be validated.
PICK Tool	The software product which Team404 is tasked with constructing for the clients.
Red Team	The title given to the attackers in the simulated cyber-attack; The team that will attack the system the blue team will defend.
Sanitized Logs/Sanitized Log Entries	A log entry that has gone through validation.
Significant Event	Logs that are important to the overall cyber-attack scenario between the Blue Team and the Red Team.

Team404	CS4311 Team 6, the software development team; this includes Mr. Alejandro Zamora, Mr. Eduardo Jiménez Todd, Mr. Jacob Torres, Mr. Jorge Felix, and Mr. Matt Montoya.
Timeline	A set of events, logs, that help convey the overall story of the events that took place between the Blue Team and the Red Team.
Timestamp	The time that a log was recorded in the system during the simulated cyber-attack.
User/Users	A person that interacts with the system; in PICK, the user will be an *Analyst*
Vector	The series of activities/steps an adversary executes or attempts to execute that is necessary to achieve an objective.
White Team	The title given to the observers during the simulated cyber-attack. They observe what happens between the Blue Team and the Red Team on the system.
Wildcard	Unsearchable characters, and any characters that can be searched through.
Zulu Time	The military and navigation parlance for the UTC time standard.

1.4.2. Acronyms

TERM	DEFINITION
AA	Adversarial Assessment
AKA	As Known As
CSV	Comma Separated Value.
DOD	Department of Defense.
ETL	Extract Transform Load.
GUI	Graphical User Interface
IEEE	Institute of Electrical and Electronics Engineers
JPG/JPEG	Joint Photographic Experts Group.
NLP	Natural Language Processor.
OCR	Optical Character Reader.
OS	Operating System
PDF	Portable Document Format.
PICK	PMR Insight Collective Knowledge.
PMR	Prevent, Mitigate Recover.
SDD	Software Design Document

1.4.3. Abbreviations

Admin	Administrator

1.5. Overview

This section provides an overview of the sections further described in this document.

1.5.1. Decomposition Description

Specifications on how designers and maintainers use the product specifications to define most important design entities for purposes such as figuring out which entity is responsible for which particular functions.

1.5.2. Detailed Description of Components

A detailed description of each of the components mentioned in the previous section.

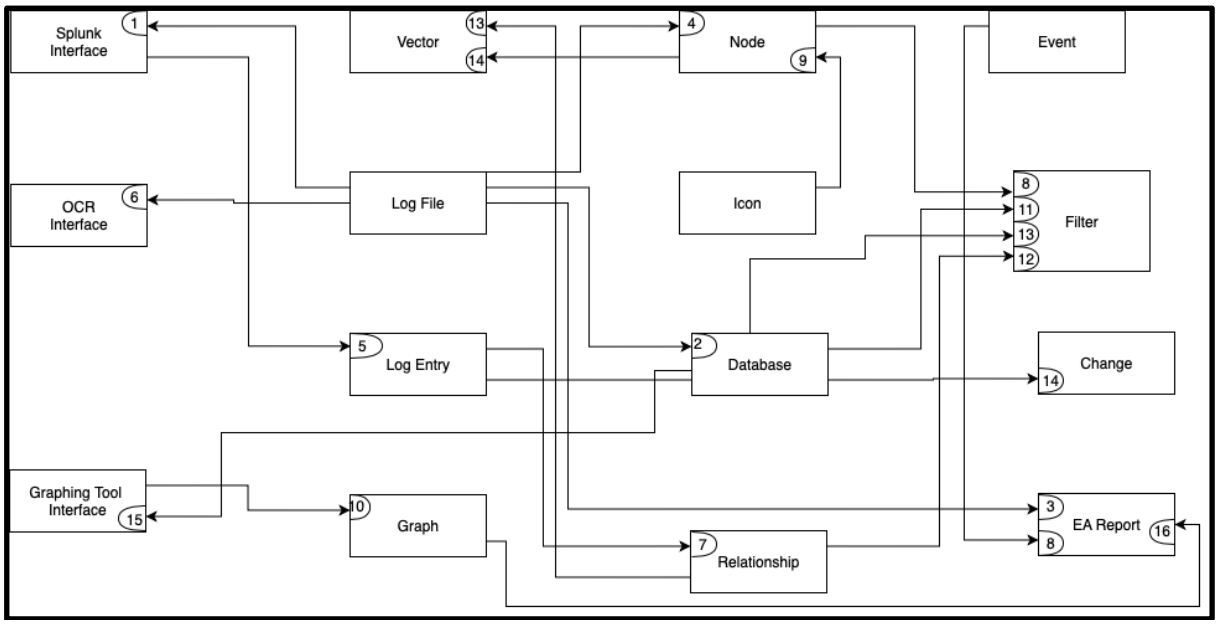
1.5.3. Database

A description of the layout of the database.

2. Decomposition Description

Section 2 contains shall introduce the System Collaboration Diagram (SCD), as well as the Subsystem and Component Descriptions (CRC Cards), as well as the dependencies of the system, for purposes such as determining which entity is responsible for specific functions and tracing requirements to design entities.

2.1. System Collaboration Diagram



2.2. Subsystem and Component Descriptions

Class: Splunk	
Superclass: None	
Subclasses: None	
Description: A third party software tool for transforming log files into normalized log entries	
Responsibilities: 1. Transform a log file into normalized log entries. 2. Store log entries in the normalized data files. 3. Process incoming data into individual activities according to the nature of the data.	**Collaborations:**
Contracts:	

Class: Maltego	
Superclass: None	
Subclasses: None	
Description: A third party software tool for graphing.	
Responsibilities: 1. Create graphs with nodes and relationships between nodes.	**Collaborations:**
Contracts:	

Class: Log File	
Superclass: None	
Subclasses: None	
Description: A computer data object that stores logs.	
Responsibilities: 1. Contain transcription with timestamps in one-minute intervals if the log file is of type “audio” or “video”. 2. Contain extracted text if the log file is of type “image” or “pdf”. 3. Contain a timestamp per line, bounded by the start date, end date, start time, and end time specified in the event configuration. 4. When a timestamp property of a previously saved event is changed, the impact of the change shall be restricted to the “not-validated” log files. 5. Store the following components: a. Log File Name b. Cleansing Status c. Validation Status d. Ingestion Status e. Acknowledgment, as well Status	**Collaborations:** 1. Will Provide log files to Splunk to transform log files into normalized log entries[Splunk (1)]. 2. Will provide log files for the event of having to create an Enforcement action report for errors in log files[Enforcement Action Report (3)]
Contracts: 1. Provide log files 2. Generate Enforcement Action Report	

Class: Significant Log Entry	
Superclass: Log Entry	
Subclasses: None	
Description: Logs that are important to the overall cyber-attack scenario between the Blue Team and the Red Team.	
Responsibilities: 1. Contain relevant information for a vector. 2. Will store the following: a. Log Entry Number b. Log Entry Timestamp c. Log Entry Content d. Host e. Source f. Source Type	**Collaborations:**
Contracts:	

Class: Node	
Superclass: None	
Subclasses: None	
Description: Visual representation of a significant event that was marked for the current vector.	
Responsibilities: 1. Represent a log entry as part of the graph 2. Will contain the following: a. Node ID b. Node Name c. Node Timestamp d. Node Description e. Log Entry Reference f. Log Creator g. Event Type h. Icon Type i. Source j. Node Visibility	**Collaborations:** 1. Will be part of at least one graph[Graph (1)] 2. Will use relationships to link to other nodes [Relationship (1)] 3. Will be displayed with an icon [Icon (3)]
Contracts: 3. Provide information to the graph. 4. Link using relationships 5. Use an Icon	

Class: Relationship	
Superclass: None	
Subclasses: None	
Description: Represent the relationship between nodes.	
Responsibilities: 1. Correlate nodes to one another. 2. Will contain the following: a. Relationship ID b. Parent ID c. Child ID d. Label	**Collaborations:**
Contracts:	

Class: Vector Database	
Superclass: None	
Subclasses: None	
Description: A database that shows cleansed logs of vectors ingested.	
Responsibilities: 1. Store the cleansed log files in permanent storage. 2. Store significant log entries in permanent storage. 3. Notify the lead when a record is pushed 4. Provide vectors to analysts. 5. Allow the lead to approve pushes. 6. Get vectors from analysts	**Collaborations:**
Contracts:	

Class: Graph
Superclass: None.

Subclasses: None.	
Description: A visual representation of a vector.	
Responsibilities: 1. A graph shall comprise of at least one node. 2. Allow the analyst to add nodes, edit nodes, delete nodes, add relationships, edit relationships, and delete relationships. 3. Create a PNG of the graph. 4. Will contain the following: a. Export Format b. Orientation c. Internal Units d. Interval e. Position of Nodes f. Position of Relationships	**Collaborations:** 1. Use Maltego to generate the graph [Maltego (1)]
Contracts: 9. Generate graphs	

Class: Vector	
Superclass: None	
Subclasses: None	
Description: A description of a significant event.	
Responsibilities: 1. Contain the following components: a. Vector Name b. Vector Description c. Log Entries	**Collaborations:**
Contracts:	

Class: OCR
Superclass: None
Subclasses: None
Description: Allows the user to make any type of document into a word-searchable document.

Responsibilities: 1. Read documents 2. Recognize character 3. Translate images to text	**Collaborations:**
Contracts:	

Class: Event Configuration	
Superclass: None	
Subclasses: None	
Description: This class enables the user to configure the events.	
Responsibilities: 1. Keep track of event names 2. Keep track of start and end times 3. Organize team folders 4. Track description of events 5. Will contain the following: a. Event Name b. Event Description c. Event Start Timestamp d. Event End Timestamp e. Root Directory f. Lead g. Lead's IP Address h. Connection Established	**Collaborations:**
Contracts:	

Class: Enforcement Action Report	
Superclass: None	
Subclasses: None	
Description: Handles errors when anomalous events happen.	
Responsibilities: 1. Track error codes 2. Display error code 3. Track description of error codes 4. Track the state of the system 5. Will contain the following:	**Collaborations:**

a. Line Number b. Error Message	
Contracts:	

Class: Icon	
Superclass: None	
Subclasses: None	
Description: An image that helps represent what the event is in the graph.	
Responsibilities: 1. Allow the change of icon 2. Display as part of a node 3. Will contain the following: a. Icon Name b. File Path	**Collaborations:**
Contracts:	

Class: Directory	
Superclass: None	
Subclasses: None	
Description: Path and structure of the log files.	
Responsibilities: 1. Perform structure check when data ingestion starts. 2. Generate root directory structure error if it fails to contain three folders or if the folder names specified in the event configuration don't match. 3. Store the log files in their corresponding folder. 4. Will contain the following: a. Red Team Folder b. Blue Team Folder c. White Team Folder	**Collaborations:**
Contracts:	

Class: Search	
Superclass: None	
Subclasses: None	
Description: Gives the user to search for specific criteria	
Responsibilities: 1. When a search operation is complete, the system shall return a result that matches the searched keyword with the searched keyword highlighted in the search result. 2. Perform logical searching. 3. Perform wildcard searching.	**Collaborations:**
Contracts:	

2.3. Dependencies

- Kali Linux.
- Python 3.6.7 or later
- Pip
- Splunk

3. Detailed Description of Component Splunk

3.1. Component Description

Splunk is a software to search, monitor and analyze machine-generated big data of applications, systems, and IT infrastructure through a web interface. Splunk captures, indexes and correlates in real-time, storing everything in a repository where you search to generate graphs, alerts, and panels easily definable by the user. The objective of Splunk is to make the data accessible to the whole organization, allowing the identification of patterns.

3.2. Class Description Splunk

Description: Splunk is third-party software that makes machine data accessible across an organization by identifying data patterns, providing metrics, diagnosing problems, and providing intelligence for business operations.

Superclass: N/A

Private Responsibilities:

- Transform a log file into normalized log entries.
- Store log entries in the normalized data files.
- Process incoming data into individual activities according to the nature of the data.

3.2.1. Contract: Provide Log Files

Contract Identifier: #1

Description: Acquire all necessary Log Files from the Adversarial Assessment

load(log_files)

Responsibilities:

- Transform a log file into normalized log entries.
- Store log entries in the normalized data files.
- Process incoming data into individual activities according to the nature of the data.

Preconditions

- Log Files need to exist
- The system needs to be connected to Splunk

Postconditions

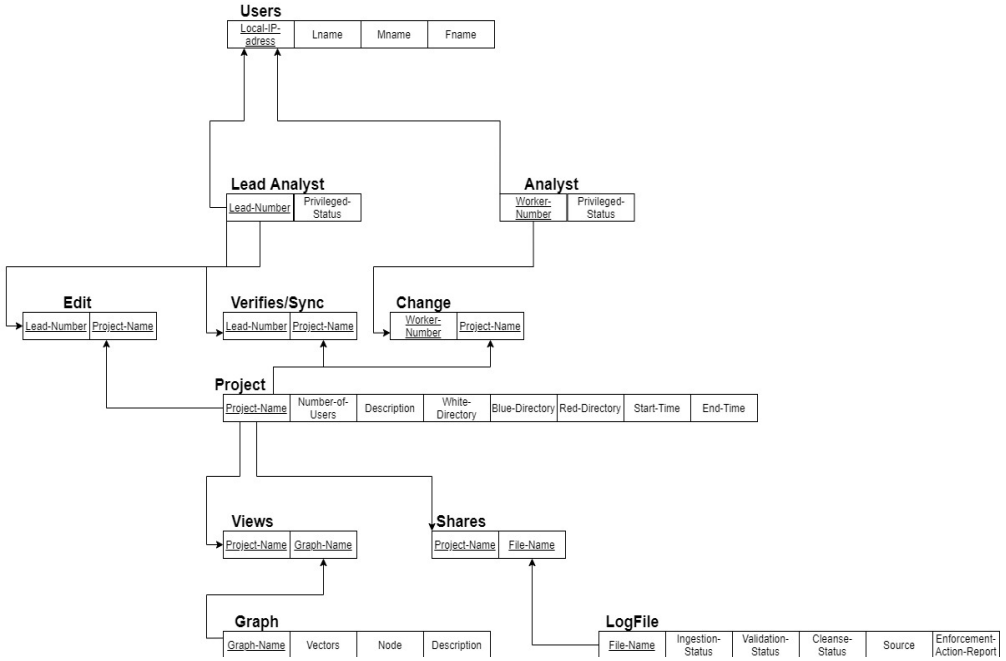
- Log Entries will be ingested to the system.

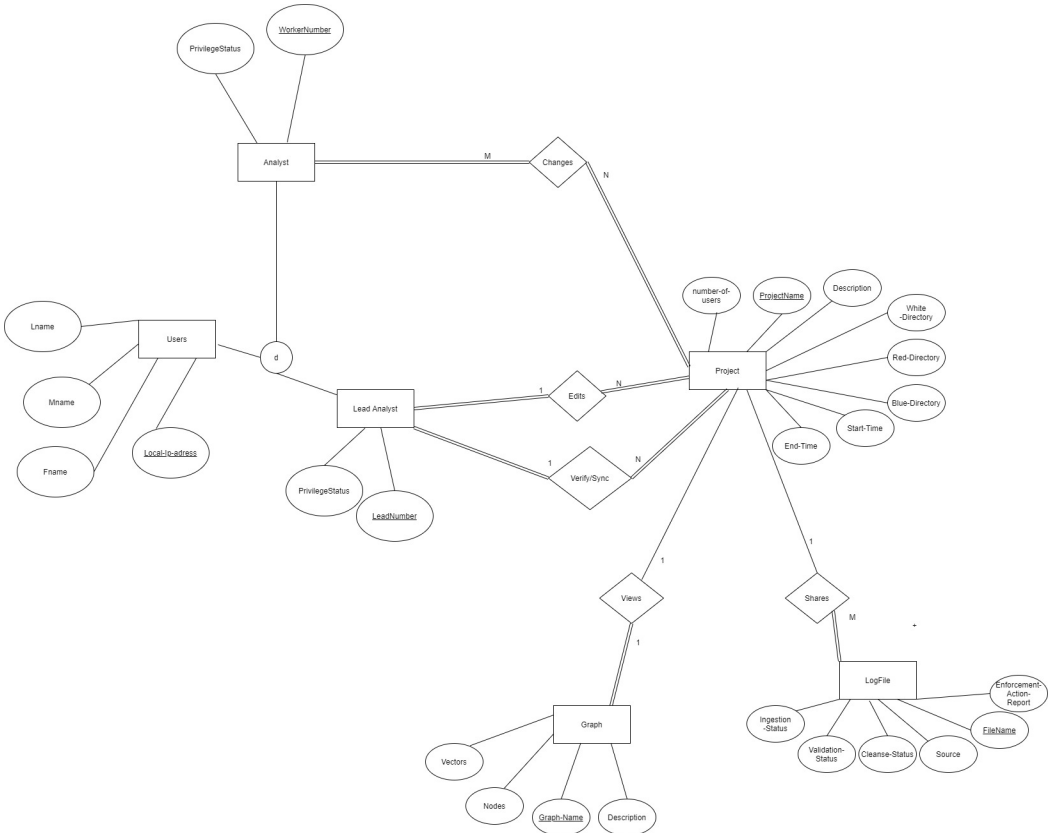
Collaborations

- N/A

4. Database

4.1. Database Schema





$