Configuration Manual

Automated Phishing Detection Framework Leveraging Integrated Threat Intelligence and Multi-UserAgent Analysis

MSc Research Project
MSc Cybersecurity

Yuvaraj Mohan
Student ID: x22200142

School of Computing
National College of Ireland

Supervisor: Raza Ul Mustafa
Industry Supervisor: Colm Gallagher

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Yuvaraj Mohan

Student ID: X22200142

Programme: MSc Cybersecurity **Year:** 2023-2024

Module: MSc Internship

Lecturer: Raza Ul Mustafa

Submission Due Date: 16-09-2024

Project Title: Automated Phishing Detection Framework Leveraging Integrated Threat Intelligence and Multi-UserAgent Analysis

Word Count: 1887 **Page Count:** 19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Yuvaraj Mohan

Date: 16-09-2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Yuvaraj Mohan
Student ID: x22200142

1 Introduction

This configuration manual is aimed at describing the process of using the Automated Phishing Detection Framework which was created as part of the MSc research work with the title: "Automated Phishing Detection Framework Leveraging Integrated Threat Intelligence and Multi-UserAgent Analysis". The framework is developed with consideration for the integration of the Tines platform with multiple threat intelligence tools for effective analysis of emails and PDF files that are suspected to contain phishing content.

The manual first outlines the prerequisites required for this framework in section 2. It is followed by section 3 which provides guidance on configuring the Tines Account. Section 4 provides the necessary information for structuring the workflow setup followed by section 5 that consists of details required for Consolidation and Reporting of the analyzed results. Section 6 provides insights to replicate the case studies conducted as part of this thesis.

2 Prerequisites

Before setting up the framework, ensure you have the following:

2.1 Tines Platform Access:

- A valid account on the Tines platform with required permissions.

2.2 Threat Intelligence Tools:

API keys for the following services:

- URLScan.io for dynamic URL analysis(*API Documentation - urlscan.io*, no date).
- VirusTotal for multi-engine file and URL scanning (*Virustotal.com*, 2024).
- EmailRep.io for real-time email reputation scoring(*Simple Email Reputation*, no date).
- Hybrid Analysis for sandboxing and behavioral analysis of PDFs(*Free Automated Malware Analysis Service - powered by Falcon Sandbox*, no date).

2.3 Programming Knowledge:

- Basic understanding of Python for scripting.
- Knowledge of RESTful APIs for service integrations.

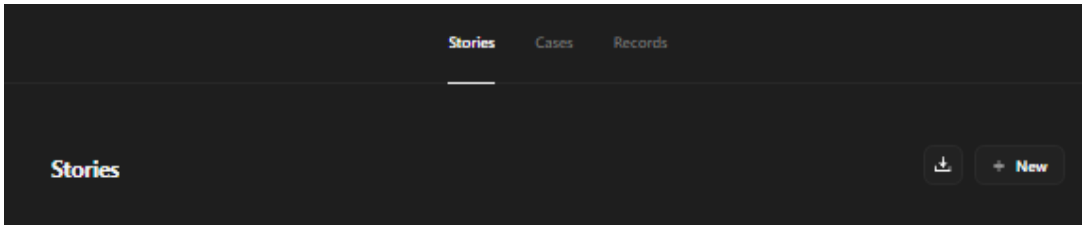
2.4 Environment Requirements:

- A secure environment for running the Tines platform.
- Internet access for API calls to the external threat intelligence services.

3 Configuring Tines Account

3.1 Creating a Tines Account

1. Open the Tines platform website(*Sign up | Tines*, no date) and create an user account.
2. Verify your email and log in to the platform.
3. Set up access control for the tenant to define roles and responsibilities(*User administration | Docs | Tines*, no date).
4. Navigate to the "Story" section to create a new story.

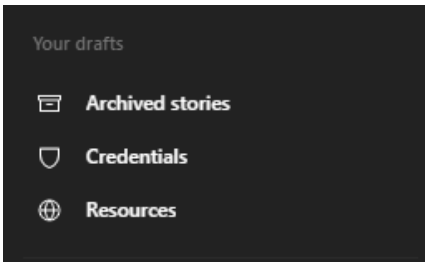


3.2 Configuring API Keys

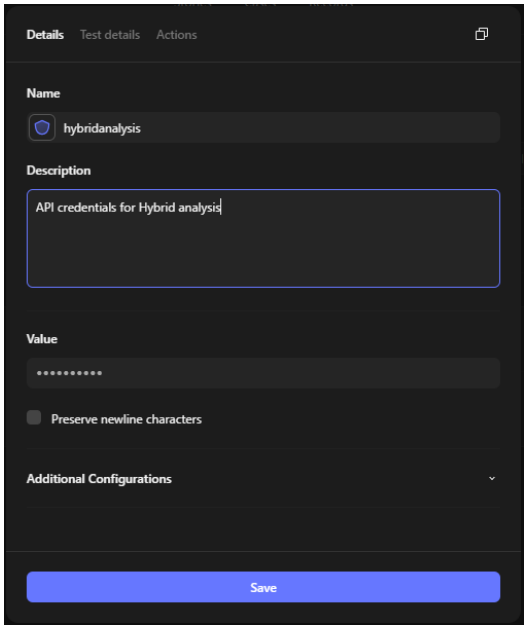
1. **Storing API Keys Securely:**

   In order to store API keys for the threat intelligence services,

   a) Go to the "Credentials" section in Tines.

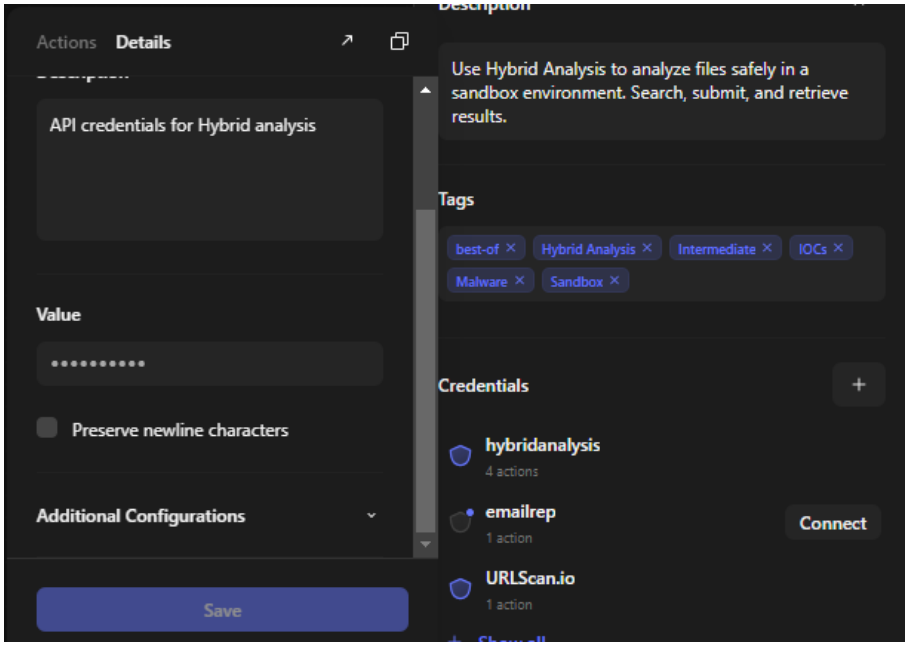


   b) Create and Add credentials for the following services:
      i. URLScan.io
      ii. VirusTotal
      iii. EmailRep.io
      iv. Hybrid Analysis



2. **API Key Configuration:**

   a) Ensure that each API key is correctly assigned to its corresponding service.
   b) Store the API keys as using encrypted storage to secure them.

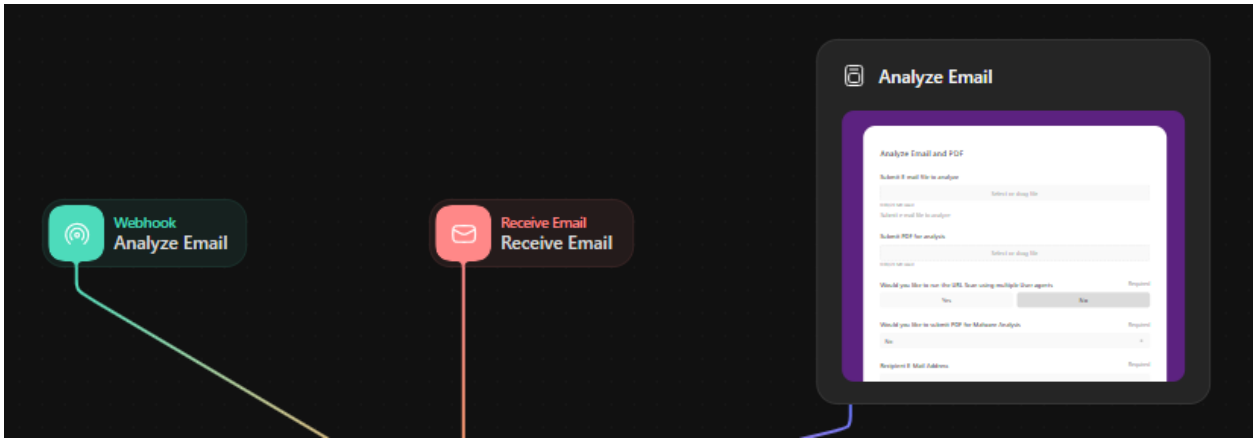


4. Workflow Setup

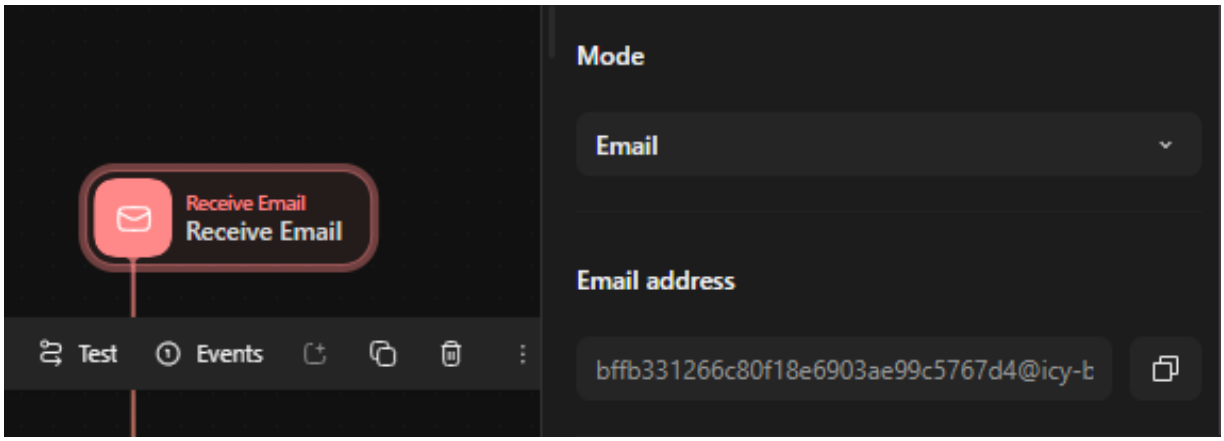
In this section, we will look at how the implementation of the workflows are configured,

4.1 Input Configuration:

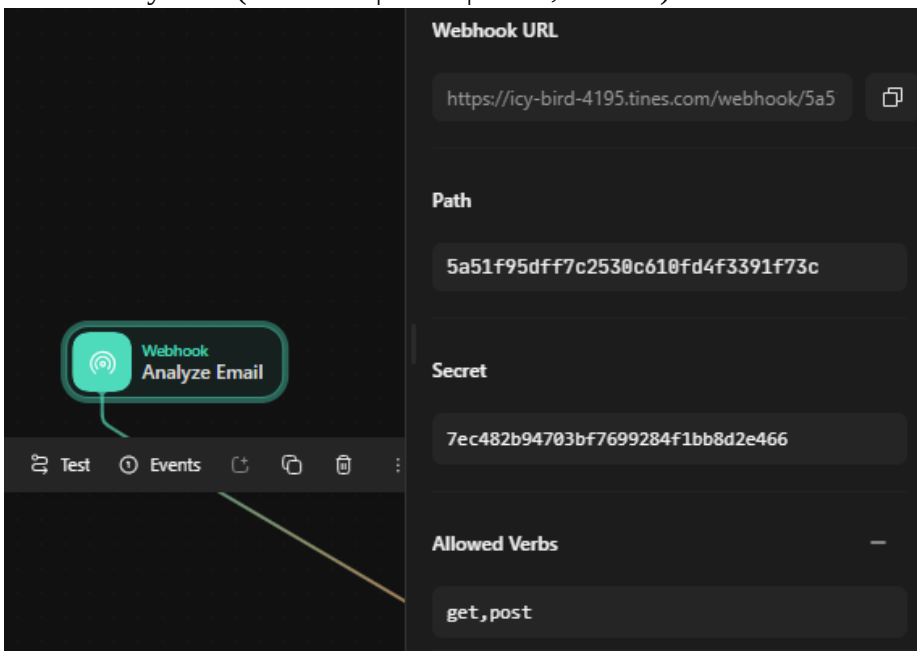
This section provides the input configuration of both email and pdf analysis workflows.



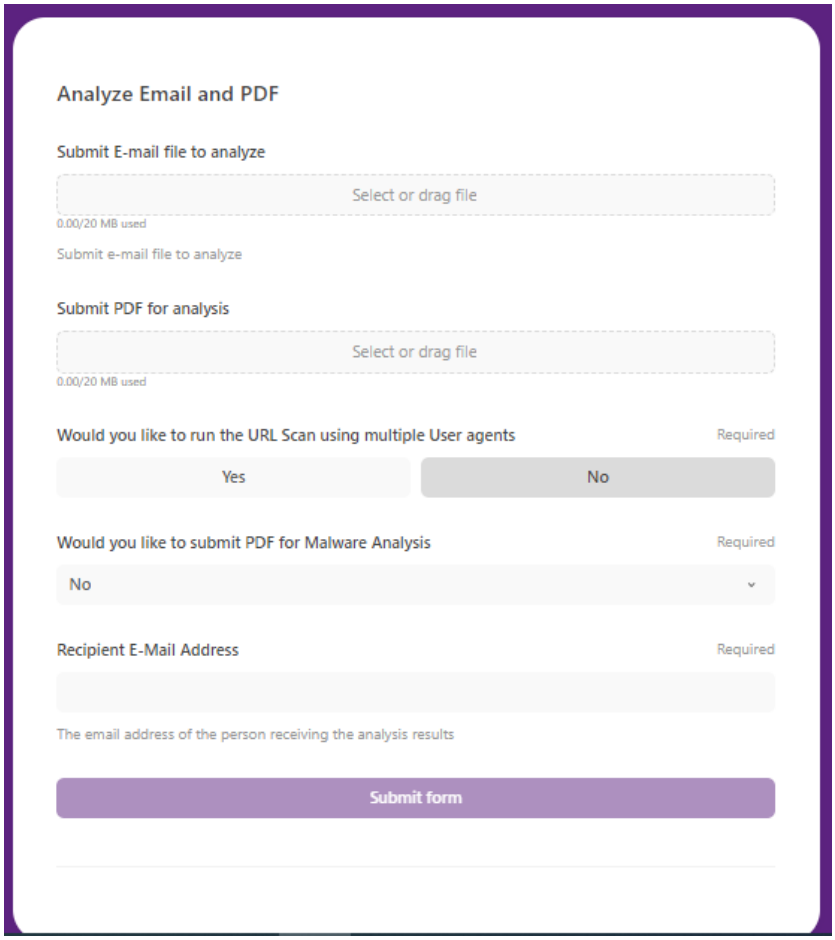
a.) Receive Email Action: In this tines action we configure an IMAP action which allows users to send email to a designated mailbox that automatically forwards the email for analysis(*Receive Email | Docs | Tines*, no date).



b.) Webhook Integration: Webhook is configured in a way to receive real-time data from external systems(*Webhook | Docs | Tines*, no date).



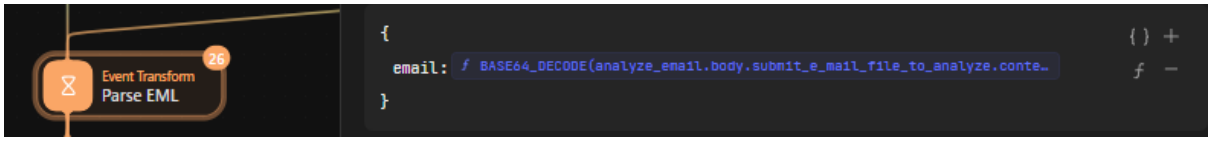
c.) ***Manual Submission:*** It allows users to upload emails directly through the UI.



4.2 Email Analysis Workflow:

a.) Email Parsing:

Setup: Configure a event transform action to use BASE64 decoding and a message parsing function to extract email contents.

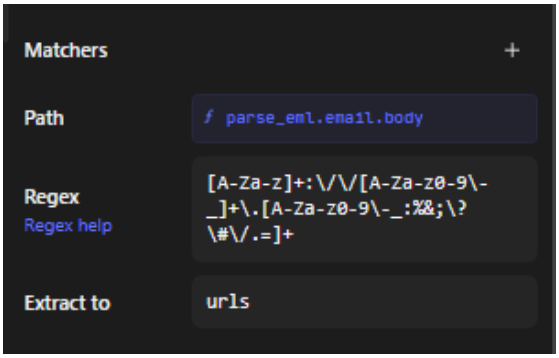


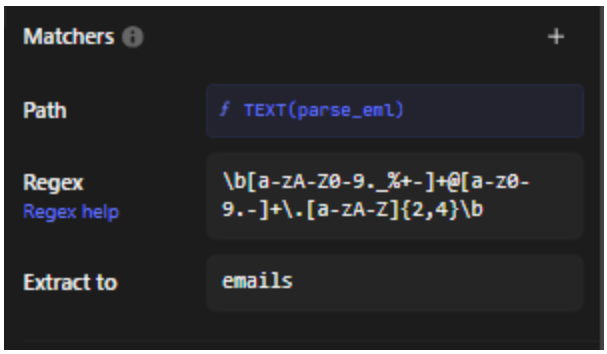
Output: The parsed email is then converted into a structured JSON format for further analysis in the workflow.

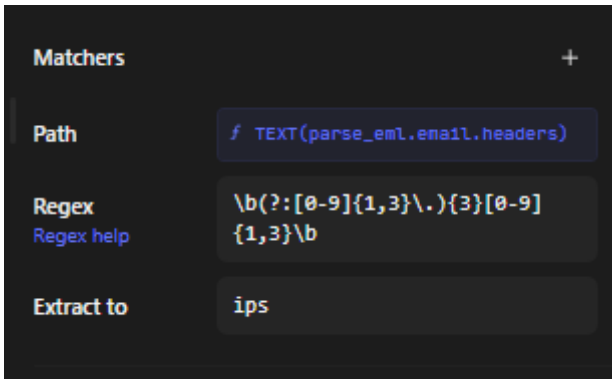

```
"parse_eml": {
    "email": {
        "message_id": null,
        "subject": "Purchase Order",
        "from": null,
        "to": [
            "tanjina@mpower-social.com"
        ],
        "cc": [ … ],
        "date": "2024-08-28T10:43:49+00:00",
        "headers": { … },
        "body": "<html xmlns:v=\"urn:schemas-microsoft-com:vml\" xmlns:o=\"urn:schemas-microsoft-com:office:office\" xmlns:w=\"urn:schemas-microsoft-com:o…,
        "attachments": [
            {
                "filename": "TestFile - DocSign238185.pdf",
                "content_type": "application/octet-stream",
                "guid": "be45e42a-56ac-4fdd-8178-57e81cb6111a",
                "md5": "df34aaf231cc8209c6511d5116986e78",
                "sha256": "6b99cb9d989e8279a34a02d402d5dd6c07ada99368eecea19fa739eac495e4ff",
                "sizeinbytes": 24669,
                "base64encodedcontents": "JV8ERi0xLjQKMSAwIG9iago8PAovVGl0bGUgKP7/kQovQ3JlYXRvciAo/v8AdwBrAGgAdABtAGwAdABvAHAAZABmACAAMAAuADEAMgAuADYpCi9… ,
                "path": "TestFile - DocSign238185.pdf"
            }
```

b.) IOC Extraction:

Regular Expressions (Regex): Configure regex patterns to identify and extract URLs, IP addresses, file hashes, and email addresses from the parsed content.

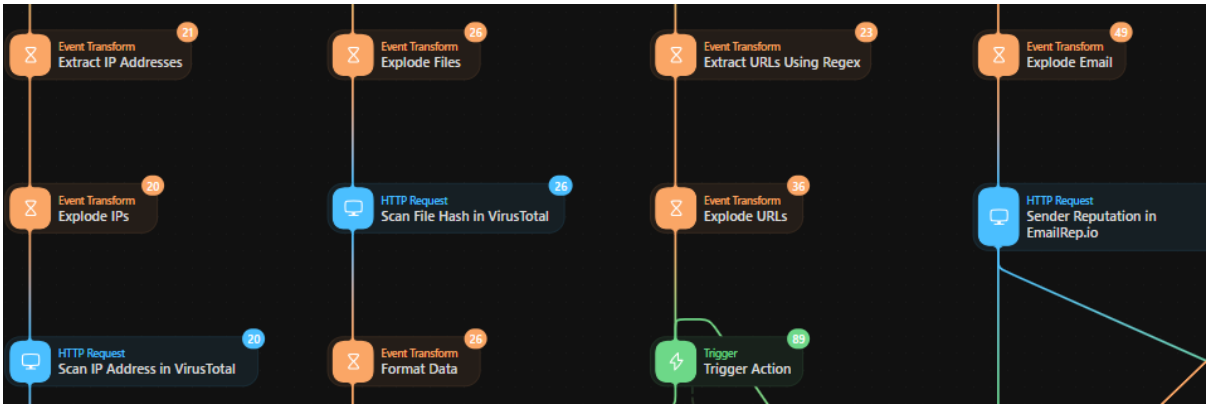






Integration with Tools: Configure to forward the extracted IOCs to their respective workflows:

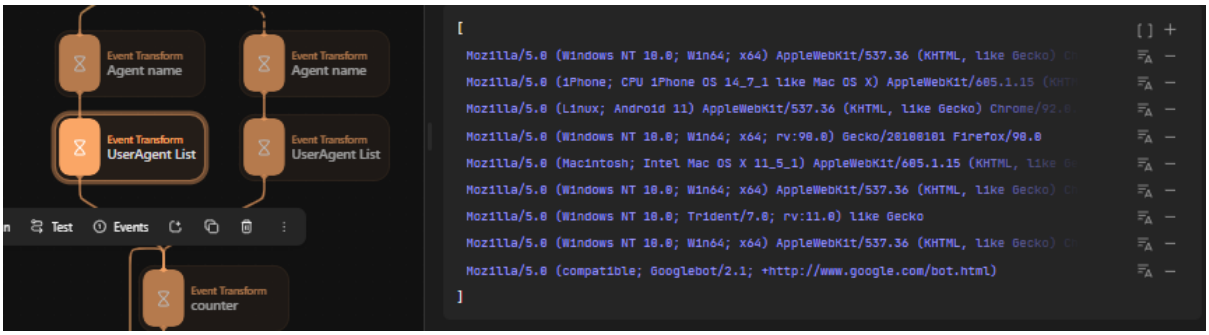
a. URLs to **URLScan.io**
b. IP addresses and file hashes to **VirusTotal**
c. Email addresses to **EmailRep.io**



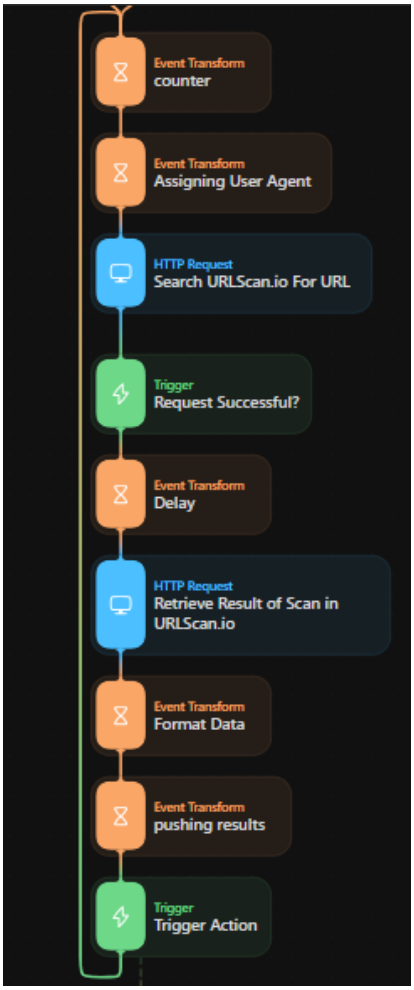
c.) Dynamic Analysis:

i) URLScan.io: Simulate and iterate through multiple user agents to analyse the URLs for cloaking or evasion techniques.

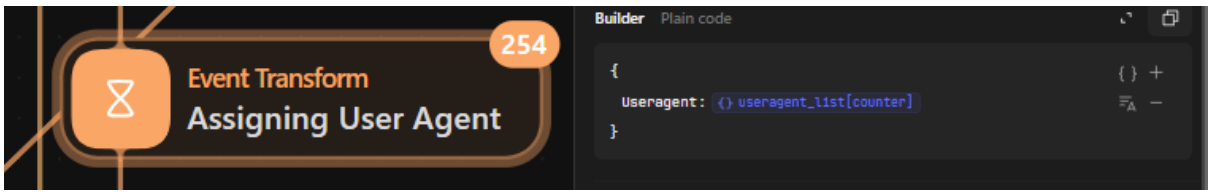
Step 1: Create an array for multiuser agent list using Tines Event Transform action.



Step 2: Create a loop structure similar to the image to loop through the user agents and retrieve results.



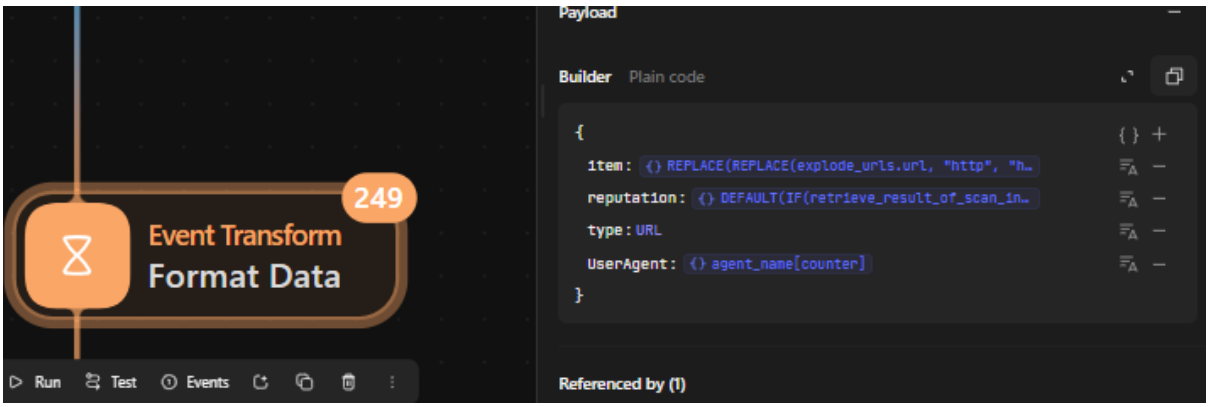
Step 3: For Each loop assign a user agent based on the counter value. This counter value acts as index for the array created in step 1.



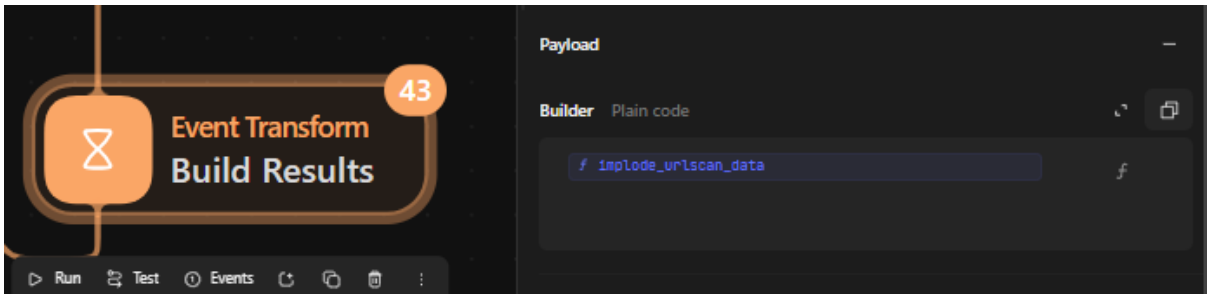
Step 4: Configure a HTTP request action , pass API key in the header and assign the counter indexed user agent as the payload for the attribute “customagent” and pass the URL fetched using regex.

Step 5: A successful request to urlscan.io will return the results similar to this image.

Step 6: Format the data and populate the results returned from the response for each iteration

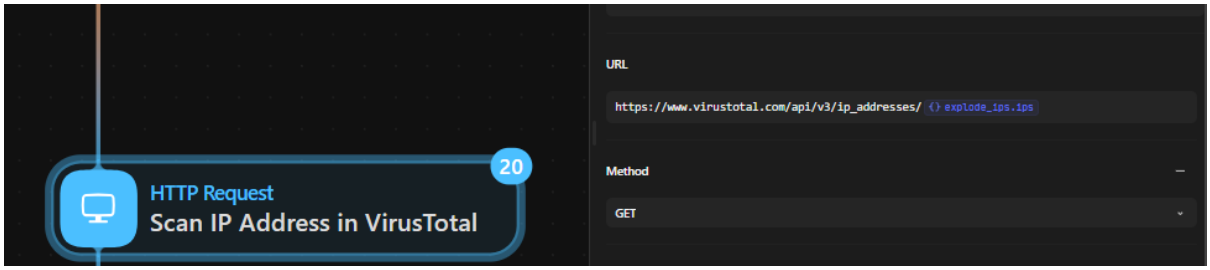


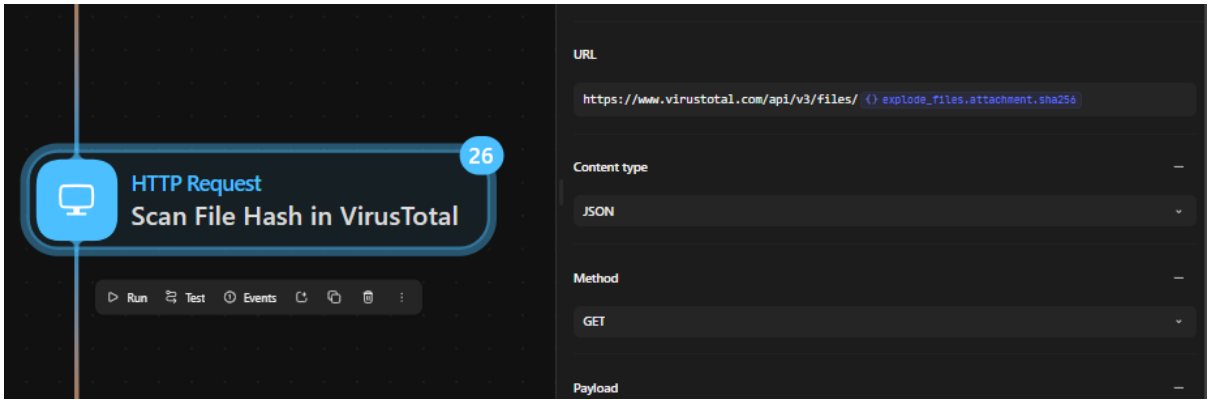
Step 7: The formatted results is then pushed into the build results where it consolidates the data from each iteration.



ii) VirusTotal: Employs both signature-based and heuristic analysis for IPs and file attachments.

Step 1: Create two HTTP request actions, assign API keys in the headers and feed the IPs and file hashes acquired using the regex as payload in the URL of the request.



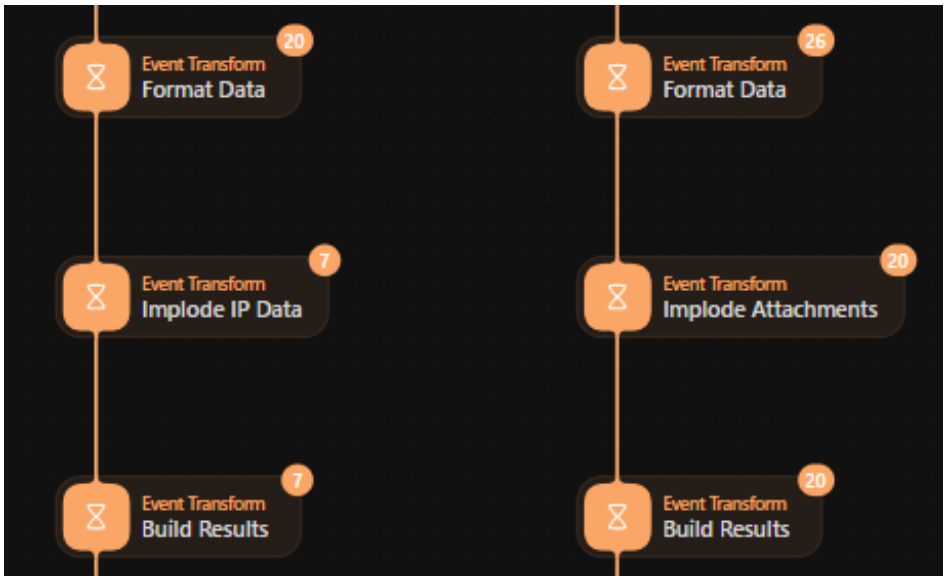


Step 2: The results from the HTTP request


```
"scan_file_hash_in_virustotal": {
    "body": {
        "data": {
            "id": "46bac0f97b2679713123b00fac0af874ab8297caeeee3dfe9685d9174a1a1390",
            "type": "file",
            "links": { ... },
            "attributes": {
                "type_tags": [ ... ],
                "magic": "PDF document, version 1.7, 1 pages",
                "last_analysis_stats": { ... },
                "names": [ ... ],
                "tlsh": "T140826E04990E3CCEF2A4AAD25F9A3D0DB86EB203D4C894C0357ECF07…,
                "creation_date": 1719394185,
                "meaningful_name": "a69900d92a437cbe482fc75e210fac0e.virus",
                "ssdeep": "384:y7IML0/LD5mjFLh/dInPP9rSNNKNHf0WSDRasLw74b9x6Cu6SiAq…,
                "size": 18159,
                "last_analysis_results": { ... },
                "last_analysis_date": 1720139820,
                "unique_sources": 1,
                "sha256": "46bac0f97b2679713123b00fac0af874ab8297caeeee3dfe9685d917…,
                "reputation": 0,
```

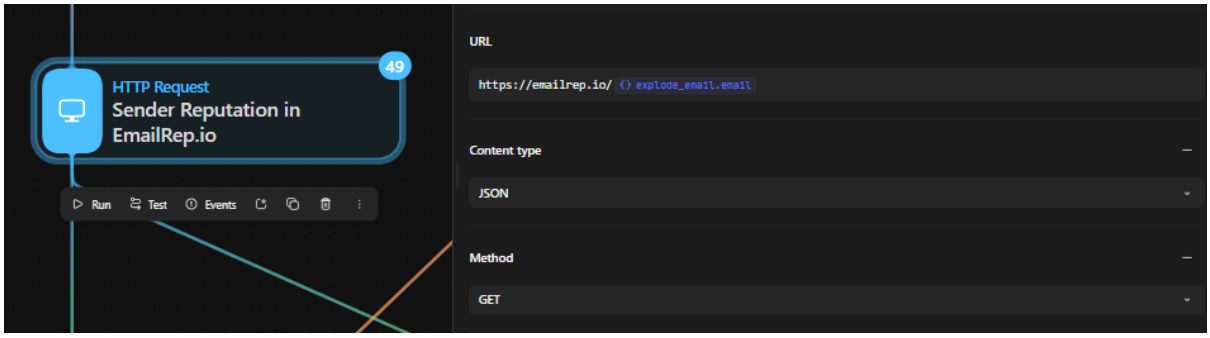
```
"scan_ip_address_in_virustotal": {
    "body": {
        "data": {
            "id": "89.144.44.2",
            "type": "ip_address",
            "links": { ... },
            "attributes": {
                "last_analysis_date": 1716531401,
                "country": "DE",
                "network": "89.144.0.0/18",
                "total_votes": { ... },
                "regional_internet_registry": "RIPE NCC",
                "reputation": 0,
```

Step 3: The response from the requests is formatted and the results are consolidated to be forwarded to reporting action.



iii) EmailRep.io: It is configured to retrieve reputation scores and analyse email addresses for potential phishing indicators.

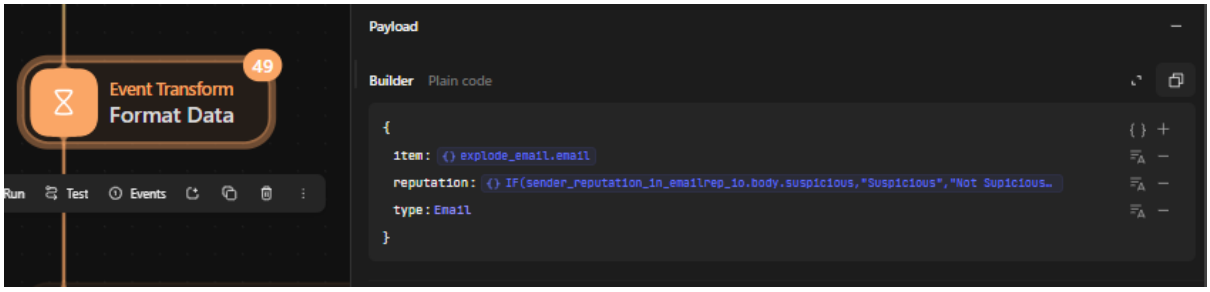
Step 1: Configure a HTTP request action and assign the API key for EmailRep.io to it. Append the email ids acquired from the regex to url.



Step 2: EmailRep.io response for the given request.


```
"sender_reputation_in_emailrep_io": {
    "body": {
        "email": "tanjina@mpower-social.com",
        "reputation": "low",
        "suspicious": true,
        "references": 0,
        "details": {
            "blacklisted": false,
            "malicious_activity": false,
            "malicious_activity_recent": false,
            "credentials_leaked": false,
            "credentials_leaked_recent": false,
            "data_breach": false,
            "first_seen": "never",
            "last_seen": "never",
```

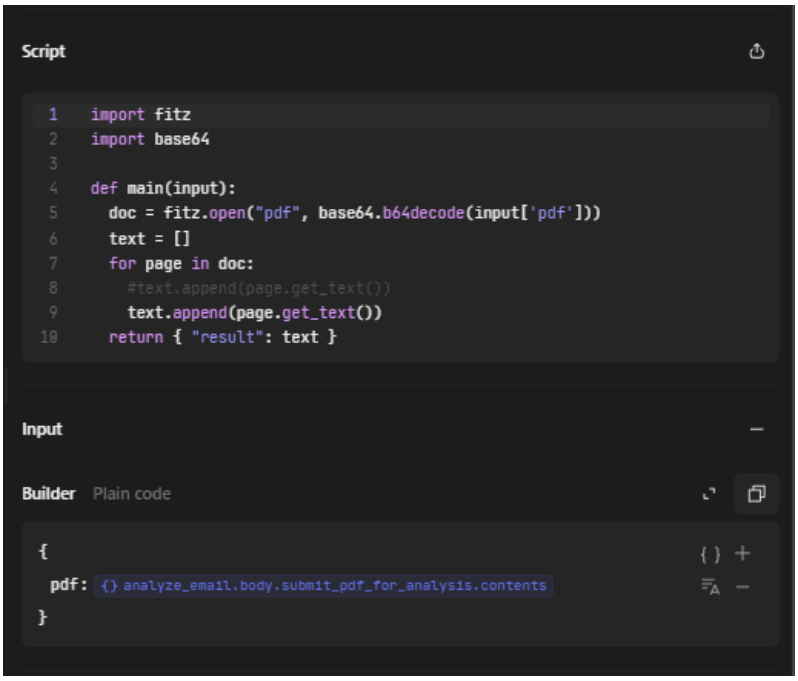
Step 3: The results are then formatted and forwarded for reporting



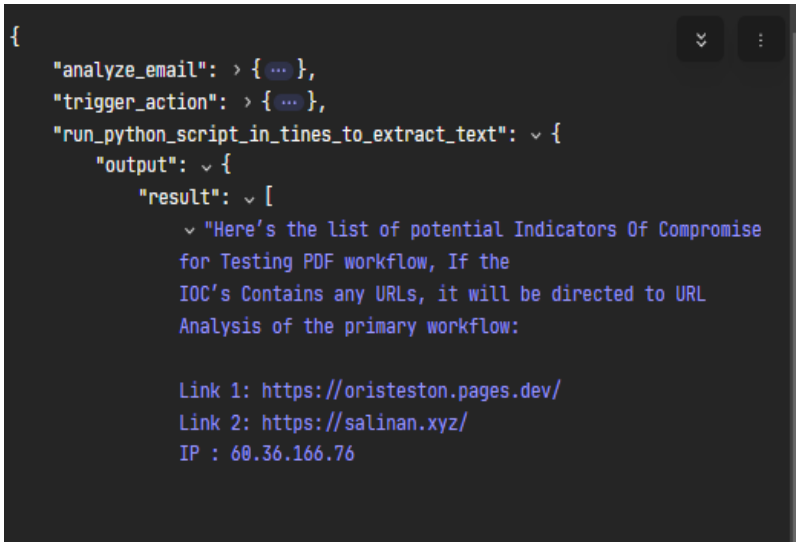
4.3 PDF Analysis Workflow:

a.) PDF Parsing:

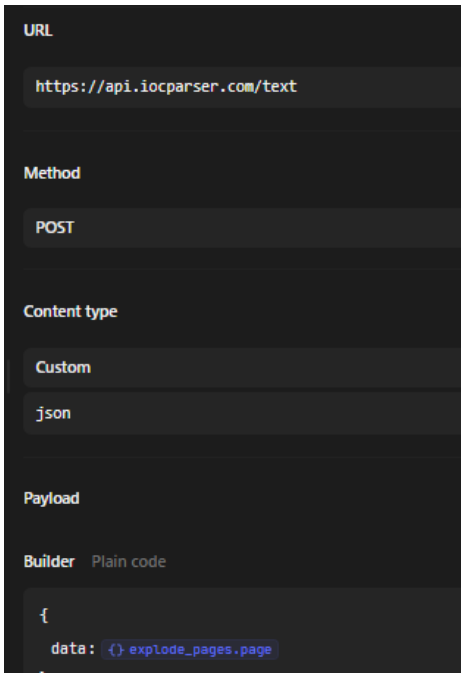
Text Extraction: Pass the pdf as the input and implement a Python script using the "fitz" module to extract text and embedded elements from the PDF.



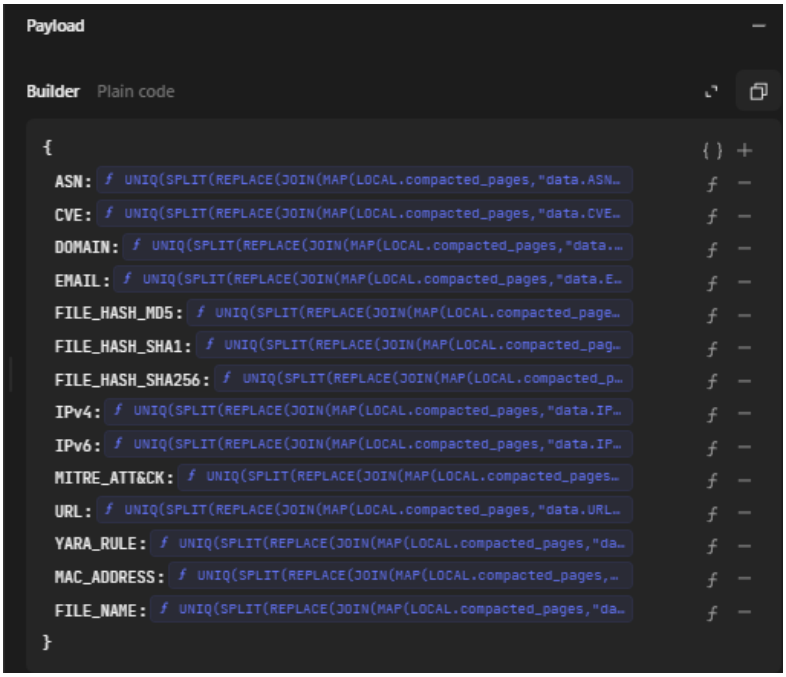
The response from the parsing lists the PDF content as text



IOC Identification: In the results from parser, use regex and iocparser.com(*IOCParser - Free IOC Extracting Service*, no date) to extract IOCs such as URLs, IP addresses, and file hashes from the parsed content.



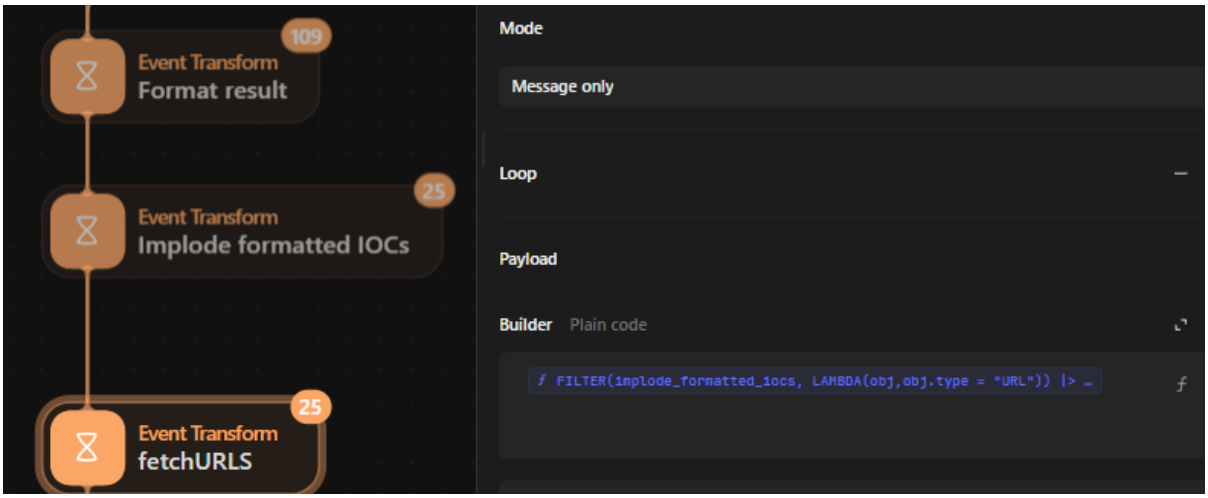
Map IOCs: The IOCs retrieved from the IOCParser.com is then Mapped in a structured JSON format for to make it display in the UI as result and also for further analysis.



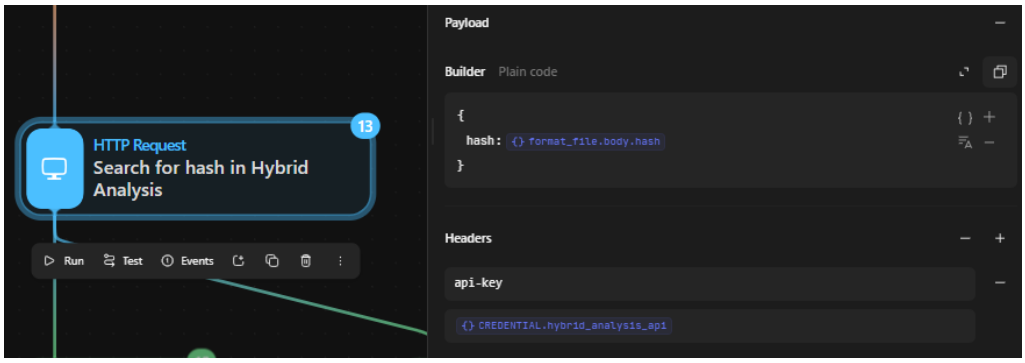
Displaying IOCs using UI: Create a Tines Page then add a table to display the results. Feed the mapped IOCs as the input to the table.

b.) Dynamic and Sandbox Analysis:

URLScan.io: Configure an event transform action to Filter and check for IOCs present in the mapped IOCs. Connect the filter function to the URLScan.io workflow to automatically forward URLs for multi-agent analysis.



Hybrid Analysis: Forward the PDF file to Hybrid Analysis for sandbox execution and detection of malicious behaviour. Create a HTTP request and add the stored API credentials in the header.



Response of the HTTP request:


```
{
    "analyze_email": > { ··· },
    "trigger_action": > { ··· },
    "format_file": > { ··· },
    "search_for_hash_in_hybrid_analysis": v {
        "body": v [
            v {
                "classification_tags": > [ ··· ],
                "tags": > [ ··· ],
                "submissions": > [ ··· ],
                "machine_learning_models": > [ ··· ],
                "crowdstrike_ai": > { ··· },
                "job_id": null,
                "environment_id": null,
                "environment_description": "Static Analysis",
                "size": 6455,
                "type": "PDF document, version 1.6",
                "type_short": > [ ··· ],
                "target_url": null,
                "state": "SUCCESS",
                "error_type": null,
```


5 Consolidation and Reporting

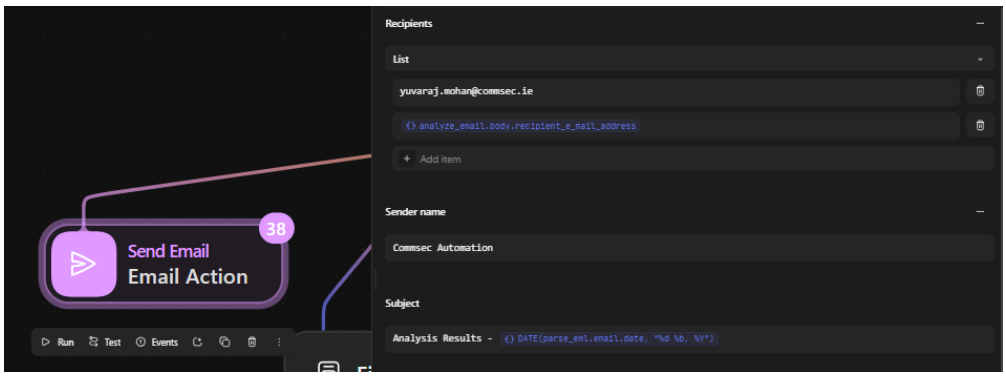
5.1 Data Consolidation:

- Use Tines' Event transformation actions to consolidate results from URLScan.io, VirusTotal, EmailRep.io, and Hybrid Analysis.
- Flatten and categorize the results for easier interpretation.

5.2 Report Generation:

- Format the consolidated data into a readable report using HTML and Tines' Send Email action.
- Configure a send email action and add the sender's mail address to the list of the recipients. Create a HTML table structure and feed the formatted results into it.
- Automatically email the report to relevant stakeholders or make it accessible via the Tines UI.

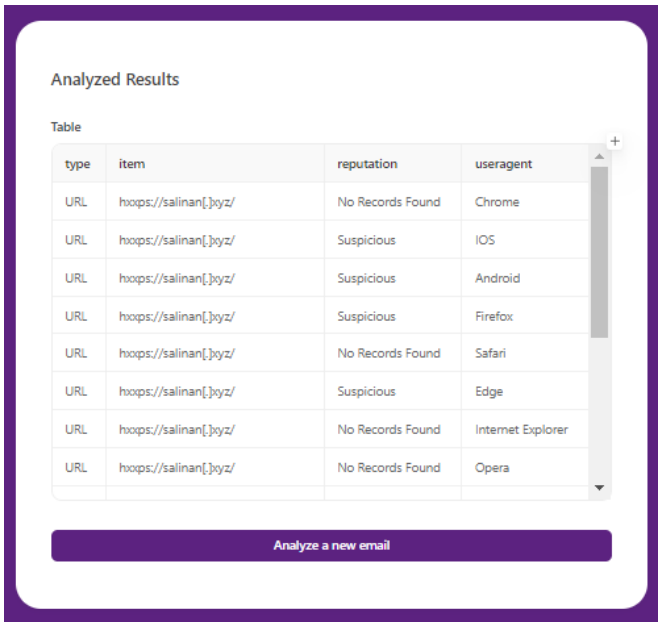
5.2.1 The email header content, email body and UI results for the email analysis are as follows,



Body

```
Hello Team,
<br><br>
Thanks for sending the email to suspicious mail box. See the attached
results from the analysis.
<br><br>
<style>
table
{
    border:1px solid #b15dde;
    border-collapse:collapse;
}
    table th {
    border:1px solid #b15dde;
    padding:18px;
    background:#8c2cd1;
    color: #313030;
}
    table td {
    border:1px solid #b15dde;
    text-align:left;
    padding:4px;
    background: #ffffff;
    color: #313030;
}
</style>

<table>
<thead>
<tr>
    <th>TYPE</th>
    <th>ARTIFACT<br></th>
    <th>REPUTATION</th>
    <th>USER AGENT  <br></th>
</tr>
```



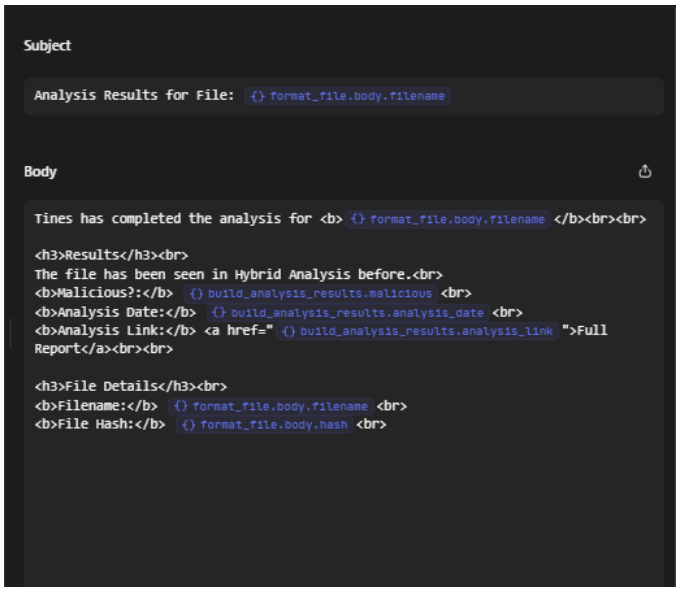
Analyzed Results

Table

type	item	reputation	useragent
URL	hxxps://salinan[.]xyz/	No Records Found	Chrome
URL	hxxps://salinan[.]xyz/	Suspicious	IOS
URL	hxxps://salinan[.]xyz/	Suspicious	Android
URL	hxxps://salinan[.]xyz/	Suspicious	Firefox
URL	hxxps://salinan[.]xyz/	No Records Found	Safari
URL	hxxps://salinan[.]xyz/	Suspicious	Edge
URL	hxxps://salinan[.]xyz/	No Records Found	Internet Explorer
URL	hxxps://salinan[.]xyz/	No Records Found	Opera

Analyze a new email

5.2.2 The email structure for Hybrid analysis is as follows,

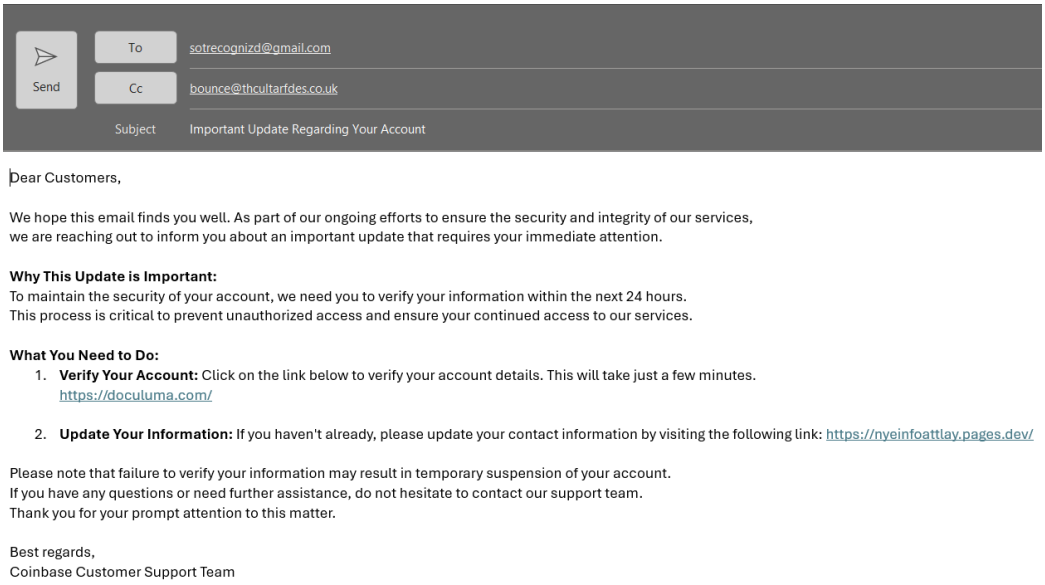


6 Case Study Implementation

The case studies provided in the thesis serve as real-world examples of the framework's capabilities. Below are the steps to replicate these cases:

6.1 Case Study 1: Malicious Email Analysis

Setup: Upload a phishing email file from available on public repositories(Corvo, 2024) or create a test email file for analysis.



Execution: Submit the test email through the webhook, manual upload or by sending the email to the designated mailbox.

Analysis: Check the extracted IOCs by analysing using the threat intelligence services.

Review Results: Assess the risk scores and detailed reports.

TYPE	ARTIFACT	REPUTATION	USER AGENT
IP Address	N/A No ipsfound	N/A	-
Email	bounce@thcultarfdes.co.uk	Suspicious	-
Email	sotrecognizd@gmail.com	Suspicious	-
URL	hxxps://doculuma[.]com/	No Records Found	Chrome
URL	hxxps://doculuma[.]com/	No Records Found	IOS
URL	hxxps://doculuma[.]com/	No Records Found	Android
URL	hxxps://doculuma[.]com/	No Records Found	Firefox
URL	hxxps://doculuma[.]com/	No Records Found	Safari
URL	hxxps://doculuma[.]com/	No Records Found	Edge
URL	hxxps://doculuma[.]com/	No Records Found	Internet Explorer
URL	hxxps://doculuma[.]com/	No Records Found	Opera
URL	hxxps://nyeinfoattlay[.]pages[.]dev/	Suspicious	Chrome
URL	hxxps://nyeinfoattlay[.]pages[.]dev/	Suspicious	IOS
URL	hxxps://nyeinfoattlay[.]pages[.]dev/	Suspicious	Android
URL	hxxps://nyeinfoattlay[.]pages[.]dev/	Suspicious	Firefox
URL	hxxps://nyeinfoattlay[.]pages[.]dev/	Suspicious	Safari
URL	hxxps://nyeinfoattlay[.]pages[.]dev/	Suspicious	Edge
URL	hxxps://nyeinfoattlay[.]pages[.]dev/	Suspicious	Internet Explorer
URL	hxxps://nyeinfoattlay[.]pages[.]dev/	Suspicious	Opera

6.2 Case Study 2: PDF File with Embedded URLs

Setup: Upload a PDF, with URLs and other IOCs embedded in it to the PDF Analysis Workflow.

Here's the list of potential Indicators Of Compromise for Testing PDF workflow, If the IOC's Contains any URLs, it will be directed to URL Analysis of the primary workflow:

Link 1: https://oristeston.pages.dev/

Link 2: https://salinan.xyz/

IP : 60.36.166.76

Execution:

- Parse and extract URLs from the PDF.
- Forward the URLs to the Email Analysis Workflow for URLScan.io analysis.

Review Results: Examine the analysis reports and identify any malicious activities.

TYPE	ARTIFACT	REPUTATION	USER AGENT
URL	hxxps://salinan[.]xyz/	No Records Found	Chrome
URL	hxxps://salinan[.]xyz/	Suspicious	IOS
URL	hxxps://salinan[.]xyz/	Suspicious	Android
URL	hxxps://salinan[.]xyz/	Suspicious	Firefox
URL	hxxps://salinan[.]xyz/	No Records Found	Safari
URL	hxxps://salinan[.]xyz/	Suspicious	Edge
URL	hxxps://salinan[.]xyz/	No Records Found	Internet Explorer
URL	hxxps://salinan[.]xyz/	No Records Found	Opera
URL	hxxps://oristeston[.]pages[.]dev/	Suspicious	Chrome
URL	hxxps://oristeston[.]pages[.]dev/	Suspicious	IOS
URL	hxxps://oristeston[.]pages[.]dev/	Suspicious	Android
URL	hxxps://oristeston[.]pages[.]dev/	Suspicious	Firefox
URL	hxxps://oristeston[.]pages[.]dev/	Suspicious	Safari
URL	hxxps://oristeston[.]pages[.]dev/	Suspicious	Edge
URL	hxxps://oristeston[.]pages[.]dev/	Suspicious	Internet Explorer
URL	hxxps://oristeston[.]pages[.]dev/	Suspicious	Opera

6.3 Case Study 3: Malicious PDF Analysis

Setup: Upload a test PDF containing EICAR test files(‘Download Anti Malware Testfile’, no date) or similar for analysis.

Execution:

- Perform initial text parsing and IOC extraction.
- Send the file to Hybrid Analysis for sandbox evaluation.

Review Results: Analyse the sandbox report for any detected threats or suspicious behaviour.

References

API Documentation - urlscan.io (no date). Available at: https://urlscan.io/docs/api/ (Accessed: 31 August 2024).

Corvo (2024) ‘rf-peixoto/phishing_pot’. Available at: https://github.com/rf-peixoto/phishing_pot (Accessed: 31 August 2024).

‘Download Anti Malware Testfile’ (no date) *EICAR*. Available at: https://www.eicar.org/download-anti-malware-testfile/ (Accessed: 31 August 2024).

Free Automated Malware Analysis Service - powered by Falcon Sandbox (no date). Available at: https://www.hybrid-analysis.com/sample/6ccc423904cb5606148879106cd6bb10007ef26fa1fcb55e60c9f8a3e8521fcc (Accessed: 1 May 2024).

IOCParser - Free IOC Extracting Service (no date). Available at: https://iocparser.com/ (Accessed: 31 August 2024).

Receive Email | Docs | Tines (no date). Available at: https://www.tines.com/docs/actions/types/receive-email/ (Accessed: 14 September 2024).

Sign up | Tines (no date). Available at: https://login.tines.com/saml_idp/signup?SAMLRequest=lZJLa8MwEIT%2Fim86yY5lk6bCNpiEQiAtJX0ceglre9MIbMnVrvv497UTStNDC73uznwzrJQRdG2vy4EPdosvAxIHJRF6Ns4unaWhQ3%2BH%2FtXU%2BLDd5OLA3JOOoh7YoGVJgyVkmSQzFbKxSGHtumgYCRRN7AhGtAhWI9hYmKjfjNY9G3vmmvQ70%2FSnhQjWq1zsGqghvlSVnEOFMq3qVFYXVS3n6b5SCtNFtWhGKdGAa0sMlnOhZiqVs4VM4vtYaZXoJH4SwePY6ZivwpkI3rvWkp4iczF4qx2QIW2hQ9Jc67vyeqNHoYava5xb%2Br89vXfsateKIpvU%2BtjOF%2F%2B%2BXYcMDTBk0TkmO73ZzRi7Xt261tQfQdm27m3pERhzwX5AEVw53wH%2FXjQO4%2BPENHJ%2FlGrswLRl03gkElFxSv35OYpP&RelayState=eyJyZWRpcmVjdF91cmwiOiJodHRwcyUzQSUyRiUyRnBhdGllbnQtc3Vuc2V0%0ALTMzMDIudGluZXMuY29tJTJGIiwicmVkaXJlY3QiOm51bGx9%0A (Accessed: 31 August 2024).

Simple Email Reputation (no date). Available at: https://emailrep.io (Accessed: 31 August 2024).

User administration | Docs | Tines (no date). Available at: https://www.tines.com/docs/admin/user-administration/ (Accessed: 31 August 2024).

Virustotal.com. (2024). Available at: https://www.virustotal.com/gui/my-apikey [Accessed 14 Sep. 2024].

Webhook | Docs | Tines (no date). Available at: https://www.tines.com/docs/actions/types/webhook/ (Accessed: 14 September 2024).