

# Visual Analytics of Control-Flow Graphs with Node-Importance Classification

Supasan Muanjit

Advisor Asst. Prof. Dr. Jakapan Suaboot

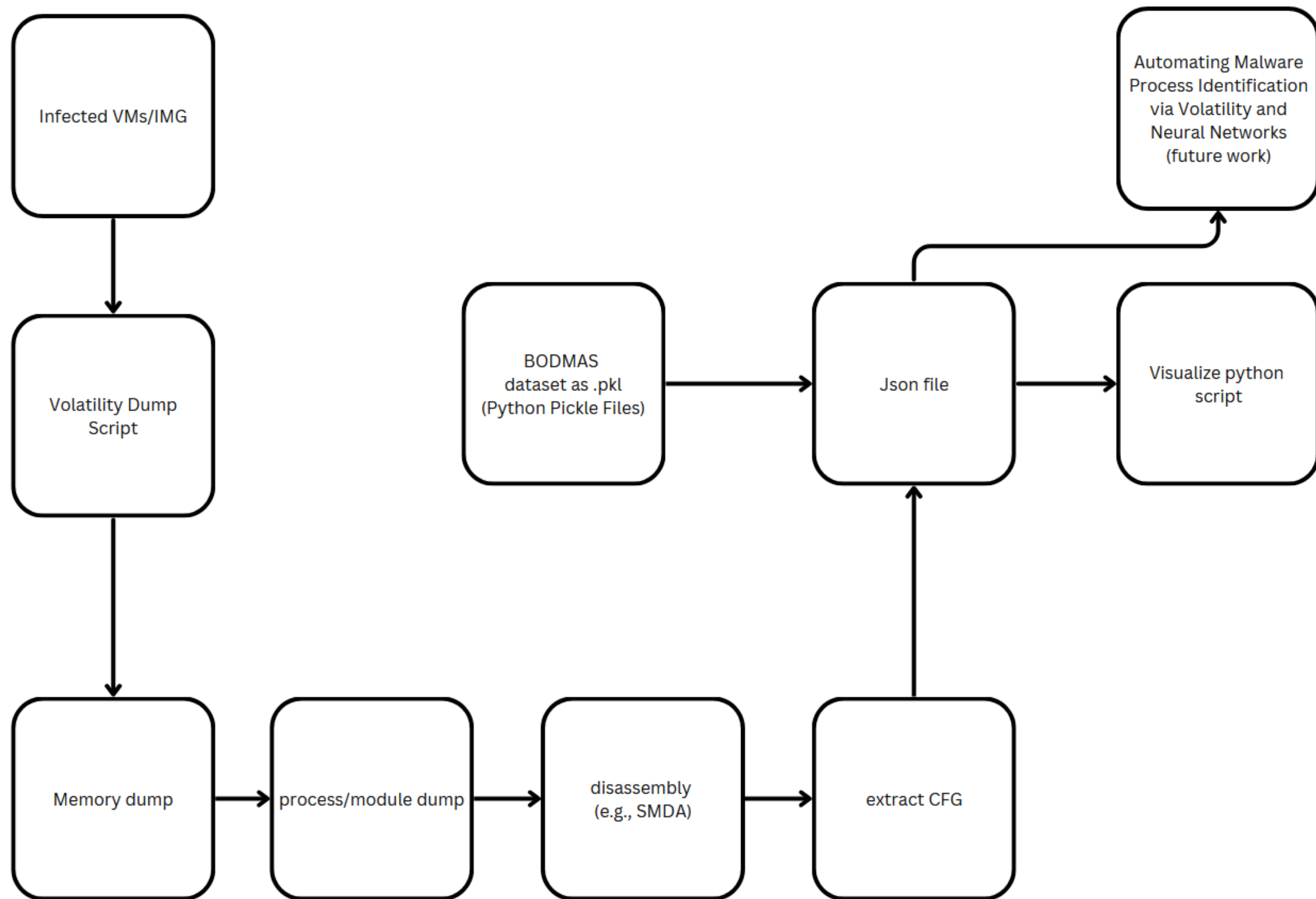
# Project Overview

- Analysis and visualization of malware behavior from pre process data
- Use of Python libraries (pandas, matplotlib, NetworkX, pyvis)
- Objective: Identify malicious process indicators visually and structurally

# Objectives

- Extract process metadata from memory dumps using Volatility Framework
- Construct Control Flow Graphs (CFGs)
- Provide interactive visualization for malware detection
- Lay foundation for machine learning integration

# Methodology



- **Memory Acquisition** Infected VMs or images are processed with Volatility to obtain raw memory dumps.
- **Process/Module Extraction** From the memory dumps, individual processes or modules suspected of being malware are isolated and dumped.
- **Disassembly** Tools such as SMDA (or similar) are used on the module dumps to disassemble executable code, allowing for static analysis.
- **CFG Extraction** Disassembled code is analyzed to extract the Control Flow Graph (CFG), which represents the execution paths within the malware's code.
- **Data Transformation** The resulting data—including CFGs—is structured into JSON files, enabling easier downstream processing and visualization.
- **Dataset Integration** Alternatively, pre-existing datasets such as BODMAS can be loaded in Python pickle (.pkl) format and converted into the same JSON structure for consistency.
- **Visualization and Automation** The JSON data can be visualized using Python scripts. The final goal is to automate malware process identification using machine learning techniques, with Volatility and neural networks suggested as future work.



# Data Exploration

# Dataset Overview

Index of /CICDataset/CIC-SGG-2024/Dataset/cfgs\_fcgs/BODMAS

Name	Last modified	Size	Description
Parent Directory	-	-	-
0a3b4a3ae8a088877a28a44e14d2a235618784f482e45b268dda47d1f34168ea.pkl	2024-10-28 18:51	256M	
0ac72e64ab8a92080428392a1297eb71b84d59a067a10e16d8fb5a93cb815979.pkl	2024-10-28 18:51	251M	
0b29e229a56fc43b6022f3cde87008ded7360a5f3e4c98c61dd9c4222a90de24.pkl	2024-10-28 18:51	250M	
0b56c37bd80c5242fca24fa9b201606824f214d8ada7fa6442eae6a6c654c779.pkl	2024-10-28 18:51	16M	
0bfedd924fd2493ef3c481a8eb2d2e4a4b11811705fb190136244c1cfac5db09.pkl	2024-10-28 18:51	253M	
0c6d6bf8036d59d2bc3882143eb86c7b0bcc3d87eee536e0c782f414a19ad712.pkl	2024-10-28 18:51	274M	
0cda7c8876136568abdb86a38b76898490ced48bfa1ac75889a7eca248155280.pkl	2024-10-28 18:51	226M	
0d7b74dc63413d1ade59fcd663f16fca483605070c546a30ead086b31441f819.pkl	2024-10-28 18:52	199M	
0da09a18dd5a491e54384bd70b7bb635b81112531a0ef3c58c26c2ae8b1704d0.pkl	2024-10-28 18:52	113M	
0ecc8c26e836b3d60a418c404138c661e2b3a49f01aa41a2604d729537a0a68c.pkl	2024-10-28 18:52	286M	

File Edit Selection View

cfgs\_fcgs > cfgs\_fcgs\_map.csv

cfgs\_fcgs > cfgs\_fcgs\_map.csv > data

label	type	dataset	hash	number_nodes	number_edges	number_weakly_connected_components	file_size
0	CFG	DikeDataset	fdb94cfe79f802c24a9e758699a241cc17df084e10638a210822aad25ccb8f50	40199	91479	102	28643641
0	CFG	DikeDataset	902a4721ddaba9f6619dd7920a9ed61d3835563adc44d207355feebd71bfa310	40904	83805	432	30162868
0	CFG	DikeDataset	270f6e7751d9866ca0334ce84e955f2789cd7a07c2a765e90539016b071fd950	48017	103681	310	35608644
0	CFG	DikeDataset	0c9baed0badb2431a9b173609d3990ccbb7c7ba4455ba0e8e8e6267d8b4ae3ef	51388	110823	358	38326412
0	CFG	DikeDataset	07615d1ba250df0b9c8772c31ec13e14b304d62ebe6258e6ae085cd8d3bacdb4	52165	106000	229	35435035
0	CFG	DikeDataset	ad3ec8fc0bde9b463ceec97c7a88ae8fba1d75caa4dde6cfadf77426fbbc57	53640	113580	192	38583629
0	CFG	DikeDataset	34e718e7c5644f147bec3af69f9fa789be1220e56a7bf827bf1c11afc9c3b0d0	55682	112395	391	42577405
0	CFG	DikeDataset	960592d2147caf51496a0531951fce5a67cb750a2719617d7bee5a869b0a0360	55700	89844	3995	44035733
0	CFG	DikeDataset	081529ae6785cf10838b4e9417c42935f10793f8acab6277b5c4821b390585e3	62658	126241	341	45254703
0	CFG	DikeDataset	b1ad1afbe24fad21bbb633124c4d21f36871a0d3cba498cfcc049eb3a358449e	64793	140865	280	45020411
0	CFG	DikeDataset	a1290e4f7bedc8c7c0c2519bd005caf7fe0210dd374490016e51ed2381731af1	70809	165852	110	51808679
0	CFG	DikeDataset	b25e024efd7212c0d84a648531a2db5c6ef16b2b85e65750e4eb99eef250547	75510	153279	207	54939178
0	CFG	DikeDataset	069d020e698f5faaa4912dcf5c2c900d431bf4182ce549c0501af48cae61450f	89949	190825	762	68738257
1	CFG	BODMAS	92522b627ca4853748872178cdd4d54c662a97bdc49d30766687a725a2c9879e	52	82	3	693242
1	CFG	BODMAS	ec14da8a4a0bd0261025bf9a4d10dde8ed574818d8c274794abb3a66f41855af	52	82	3	693244
1	CFG	BODMAS	a7cfffa10f12dbfce54fc79d21a075c3461a26fa6b986f5b570f9967720f8781	104	54	63	733948
1	CFG	BODMAS	e5113cd6508b5d6b82412b49816aff549917430f8633e97e8f8da8d8b6303f54	129	65	78	746177
1	CFG	BODMAS	26adccfe63f0e68af6b1cbe84f3571a6922a93b1cbe1e8e29fbae71bc9f49206	175	340	1	807072
1	CFG	BODMAS	5112c027bcd1e1bf290816d15121a94d4eb9a237f2140a082040b2e555579253	175	340	1	797625
1	CFG	BODMAS	26439e0097d8701794f9a1250ab4ad443c30fd2094e12181a4f50bb9c07d055b	175	340	1	807070
1	CFG	BODMAS	1b6ac6504b51ce9ed55fef22b552dd5bb65e52d67ee07919c8a0703123efd9	175	340	1	797625
1	CFG	BODMAS	0b56c37bd80c5242fca24fa9b201606824f214d8ada7fa6442eae6a6c654c779	271	450	5	16562142
1	CFG	BODMAS	65c60293319a74107ab3c923550b9192087e7c4c06ff8fbf6d51eb38b323a3fe	271	450	5	16860133
1	CFG	BODMAS	6d4d0c4687e089d2a07249e5a584585a59a3e32ebdd63fe05c162457a840f7e	609	378	306	1183341
1	CFG	BODMAS	7ed45d5f218a526479b3ef51886e428e5a3dfceb248dcb5d8f4d53d5b9b3c7f	621	394	297	1184821
1	CFG	BODMAS	8ad3338a473940a74d01a92d4ccf6c1456c654450015214f561c600a63047c63	950	1204	13	1403340

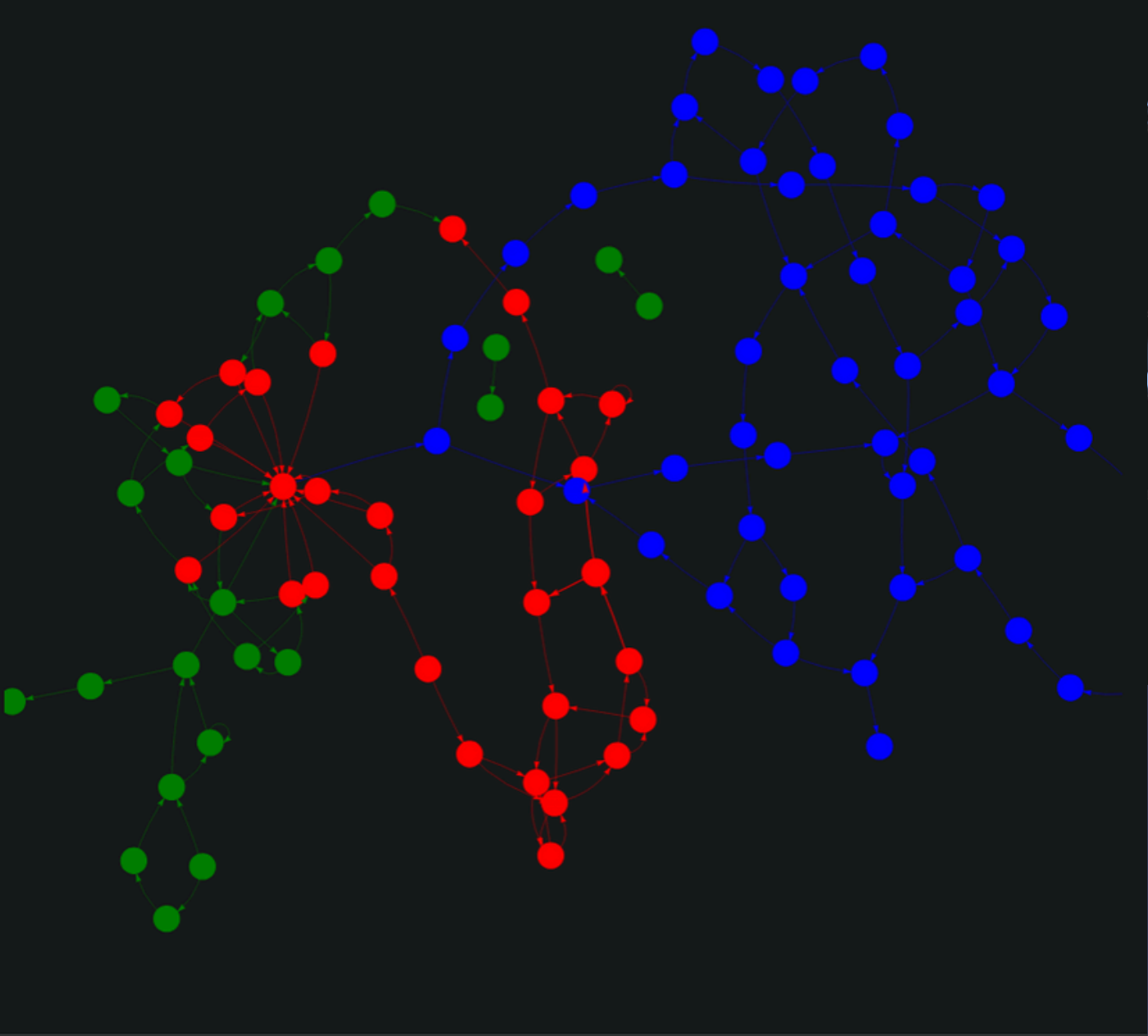
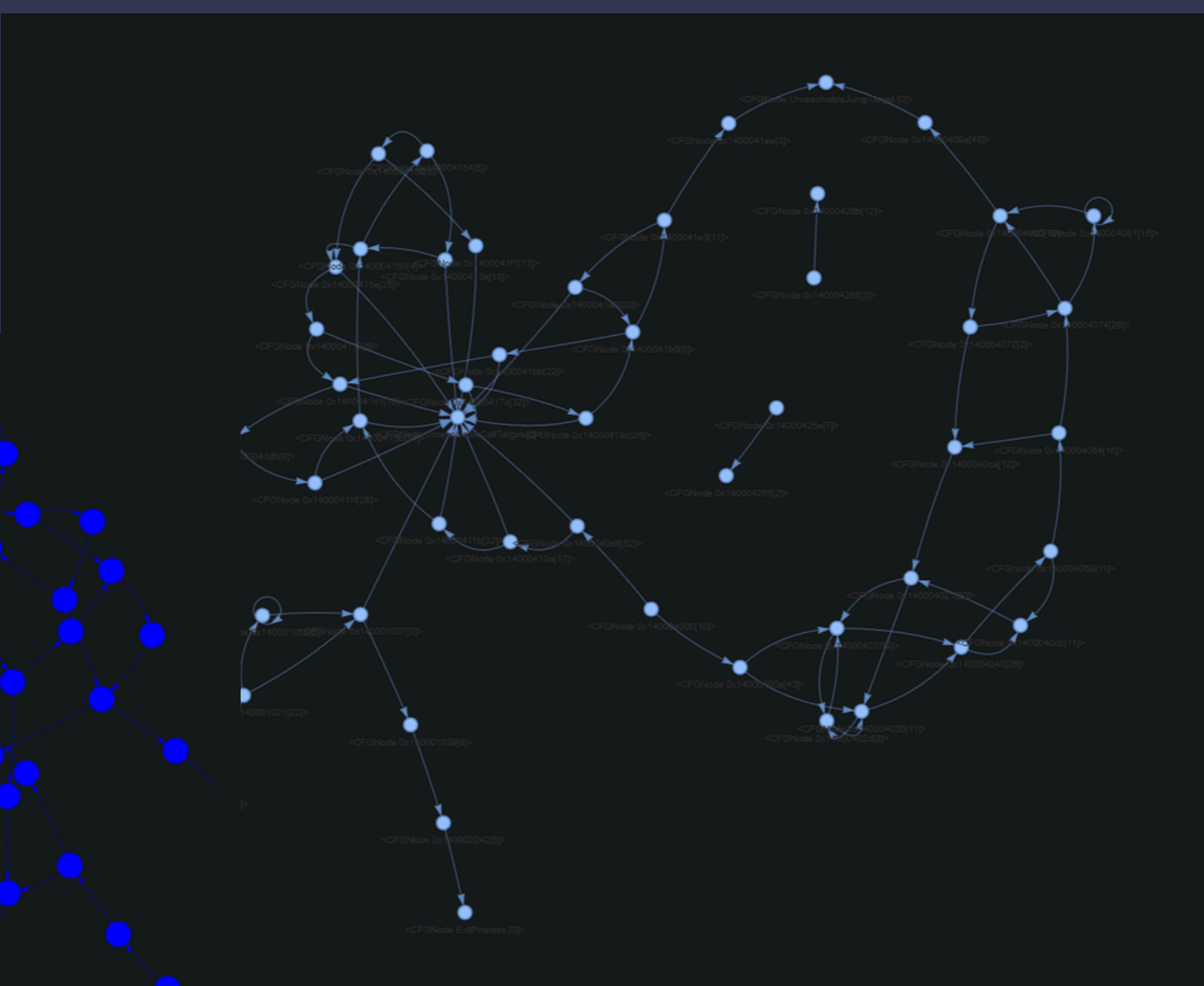
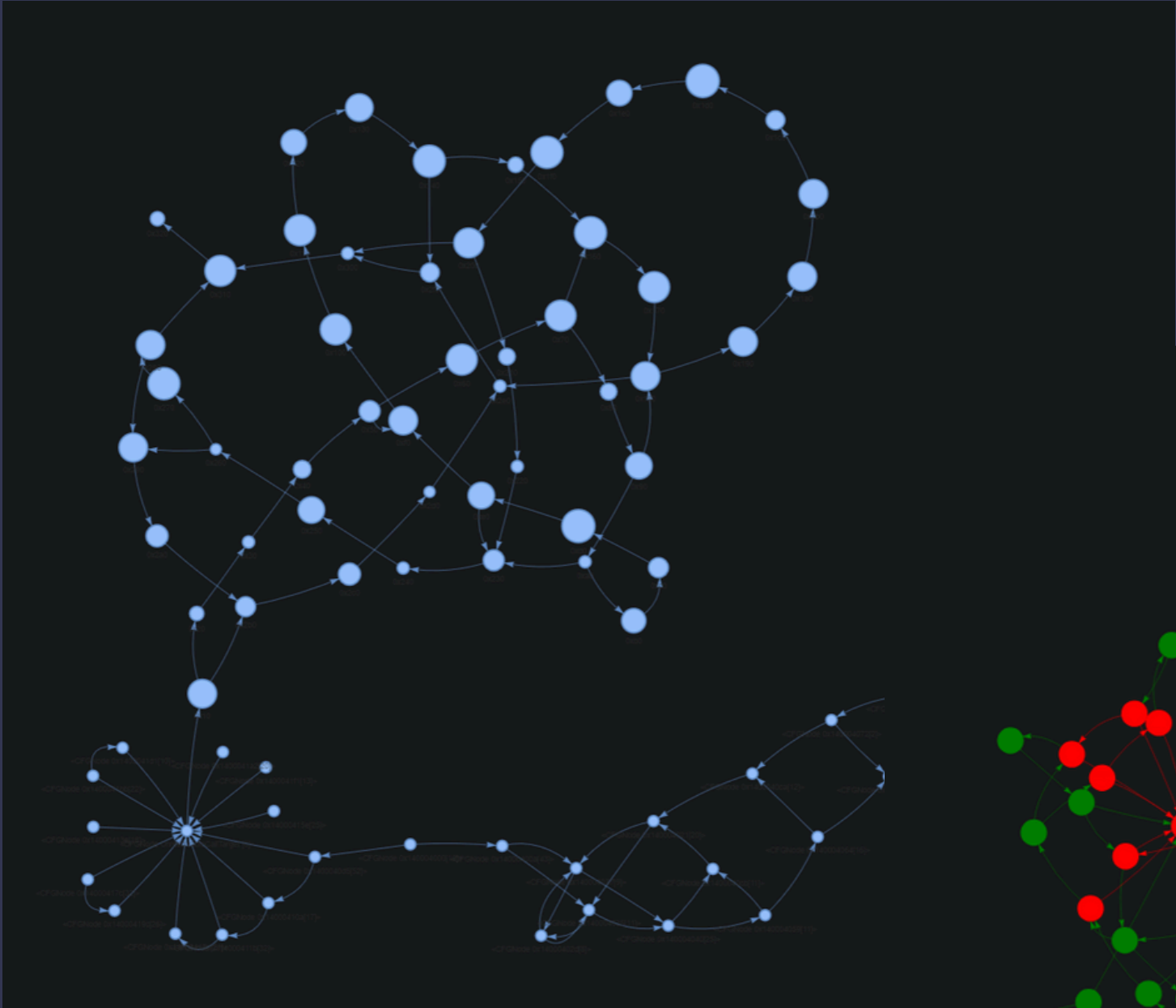
malware.json

C: > Users > Admin > Documents > Work > Project > CFGs\_y4t1 > malware.json

```
1 {
2   "directed": true,
3   "multigraph": false,
4   "graph": {},
5   "nodes": [
6     {
7       "id": "<CFGNode 0x140004000[10]>"
8     },
9     {
10      "id": "<CFGNode 0x1400040d6[52]>"
11    },
12    {
13      "id": "<CFGNode UnresolvableCallTarget [0]>"
14    },
15    {
16      "id": "<CFGNode 0x14000410a[17]>"
17    },
18    {
19      "id": "<CFGNode 0x14000411b[32]>"
20    },
21    {
22      "id": "<CFGNode 0x14000413b[21]>"
23    },
24    {
25      "id": "<CFGNode 0x140004150[4]>"
26    },
27    {
```

Dataset from  
BODMAS

# Data Virtualization



**Legend:**

- Shared nodes/edges (both files)
- Nodes/edges unique to mocking.json
- Nodes/edges unique to boombasv2.json

\*Each node’s name identifies a specific execution block or function within the analyzed process

\*The naming allows forensic analysts to trace which section of code is being executed, whether it’s benign or potentially malicious, supporting targeted investigation and cross-referencing with known malware signatures

# Demonstration

# Summary

# What has been Done

Constructed Control Flow Graphs (CFGs) representing malware execution paths using NetworkX.

Created interactive visualizations of CFGs with pyvis, showing shared and unique code paths with color coding.

Implemented basic heuristic anomaly detection to highlight suspicious nodes and edges.

# What can improve

Do web version so it can be easy to accese



# Reference

---

- [1] A. Huseinović and S. Ribić, "Virtual machine memory forensics," 2014.
- [2] A. Afreen, M. Aslam and S. Ahmed, "Analysis of Fileless Malware and its Evasive Behavior," in 2020 International Conference on Cyber Warfare and Security (ICCWS), Islamabad, Pakistan, 2020.
- [3] M. Ficco, "Malware Analysis by Combining Multiple Detectors and Observation Windows," IEEE Transactions on Computers, vol. 71, pp. 1276 - 1290, 2021.
- [4] Akash Thakar; Rakesh Singh Kunwar; Hemang Thakar; Kapil Kumar; Chintan Patel, "Enhanced Malware Detection Method Using Baseline Comparison in Memory Forensics," in 2023 IEEE 11th Region 10 Humanitarian Technology Conference (R10-HTC), Rajkot, India, 2023.



# Q&A