README: REINFORCEMENT LEARNING MALWARE INVESTIGATION DATASET

1. Dataset Overview

This dataset supports research in malware detection and investigation using reinforcement learning techniques. It contains memory dump snapshots from various Windows operating systems, both clean and infected by different malware samples. The dataset allows for detailed forensic analysis and model training to identify malware behaviours.

2. Dataset Composition

- **Operating Systems:**

  13 Windows OS versions, including:

  - Windows 7 (Professional, Starter, Ultimate, multiple editions and service packs)
  - Windows 8.1 Pro (32-bit and 64-bit)
  - Windows 10 (various updates, 32-bit and 64-bit)
  - Windows 11 ISO 22H2 (64-bit)
  - Windows Developer Evaluation (WinDev2303Eval)

- **Malware Samples:**

  13 malware types representing diverse attack behaviours:

  WannaCry, Cerber, DeriaLock, $LuckyLcoker, Dharma, SporaRansomware,

GandCrab, GoldenEye, Locky.AZ, InfinityCrypt, Win32.BlackWorm, PowerLoader, W32.MyDoom.A.

- **Files:**

  26 memory dump files total: 13 clean (uninfected) and 13 infected (one infected file per OS version).

3. Data Collection Methodology

- Experiments performed on virtual machines in an isolated, secure network environment at London Metropolitan University Digital Forensics Laboratory.
- Initial snapshots taken before infection.
- Malware was introduced to each OS environment, followed by infected snapshots.
- Memory dumps were extracted and analysed using the Volatility Memory Forensics framework
- Manual examination was performed to identify malware behaviours and system changes.

4. Data Structure

- The data structure included memory dump files in raw format.
- The metadata files provide information about the OS version, infection status, and type of malware.
- The experiment setup and malware workflow diagrams are comprehensively documented.

5. Instructions for Use

1. Identify the operating system and infection status of each memory dump via metadata.
2. Use memory forensics tools, such as Volatility to analyse the RAM dumps.
3. Extract system artefacts (process lists, network connections, loaded modules) to detect malware footprints.
4. Compare clean and infected dumps with the same OS to find behavioural differences.
5. Refer to the included malware workflow diagrams to understand typical malware behaviours and processes.
6. Use the dataset to develop, train, or validate malware detection models, especially reinforcement learning frameworks.

6. Important Notes

- The isolated setup ensures no malware spread risk.
- The data collection process and analysis are fully documented to support reproducibility.
- The dataset provides a broad range of OS versions and malware types for diverse research applications.

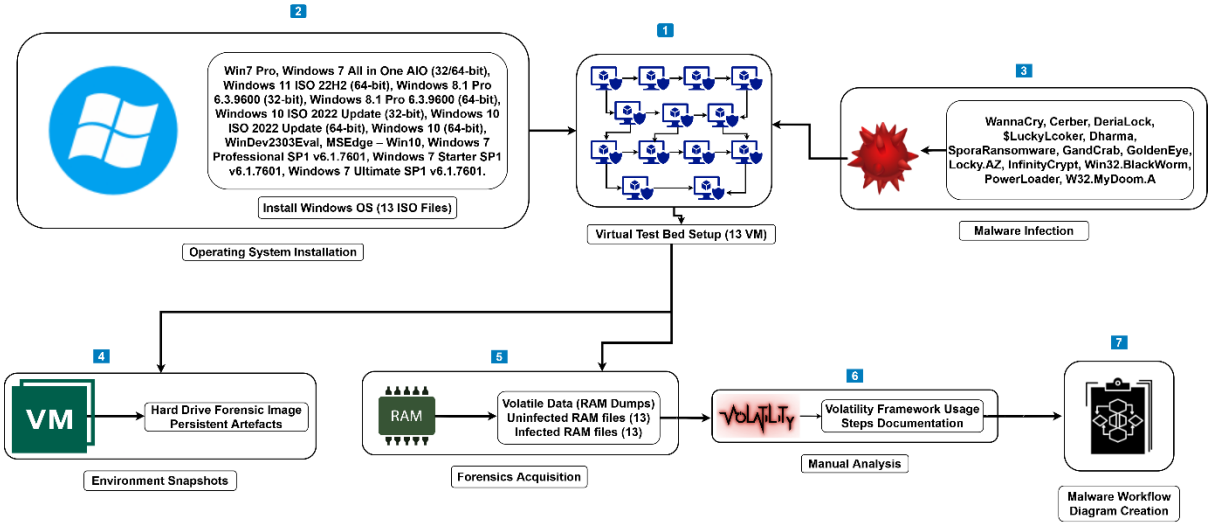

Fig 1: Malware Workflow Diagram

1. WannaCry-Win7 Pro Spanish	05/12/2023 17:14	File folder
2. Cerber - Windows 7 All in One AIO (3264-bit)	05/12/2023 17:16	File folder
3. DeriaLock - Windows 11 ISO 22H2 - 64bit	05/12/2023 17:19	File folder
4. $LuckyLcoker - Windows 8.1 Pro 6.3.9600 - 64bit	05/12/2023 17:23	File folder
5. Ramsomware - Dharma - Windows 8.1 Pro 6.3.9600 - 32bit	05/12/2023 17:26	File folder
6. Ransomeware - SporaRansomware - Windows 10 ISO 2022 Update 32bit	05/12/2023 17:29	File folder
7. Ransomeware - GandCrab - Windows 10 ISO 2022 Update-64bit	05/12/2023 17:33	File folder
8. Ransomeware - GoldenEye - Windows 10-64bit	05/12/2023 17:36	File folder
9. Ransomeware - Locky.AZ - WinDev2303Eval	05/12/2023 17:39	File folder
10. Ransomeware - InfinityCrypt - MSEdge - Win10	05/12/2023 17:43	File folder
11. Malware - Win32.BlackWorm - Windows 7 Professional SP1 v6.1.7601	05/12/2023 17:46	File folder
12. Malware - PowerLoader - Windows 7 Starter SP1 v6.1.7601	05/12/2023 17:47	File folder
13. Malware - W32.MyDoom.A. (or_and) ShadowBot - Windows 7 Ultimate SP1 v6.1.7601	05/12/2023 17:12	File folder

Figure 1: Creating a Robust Malware Dataset for the RL Framework

$LuckyLcoker - NoVirus.mem	$LuckyLcoker - WithVirus.mem	Cerber - NoVirus.mem
Cerber - WithVirus.mem	DeriaLock - NoVirus.raw	DeriaLock - WithVirus.mem
Dharma - NoVirus.mem	Dharma - WithVirus.mem	GandCrab - NoVirus.raw
GandCrab - WithVirus.raw	GoldenEye - NoVirus.raw	GoldenEye - WithVirus.raw
InfinityCrypt - NoVirus.raw	InfinityCrypt - WithVirus.raw	Locky.AZ - NoVirus.mem
Locky.AZ - WithVirus.mem	PowerLoader - NoVirus.mem	PowerLoader - WithVirus.mem
SporaRansomware -NoVirus.mem	SporaRansomware -WithVirus.mem	W32.MyDoom.A. - NoVirus.mem
W32.MyDoom.A. - WithVirus.mem	WannaCry - NoVirus.raw	WannaCry - WithVirus.raw
Win32.BlackWorm -NoVirus.mem	Win32.BlackWorm -WithVirus.mem	

Figure 2: The 26 Robust Malware Datasets for the RL Framework