Individual Project (CS3IP16)

Department of Computer Science

University of Reading

Project Initiation Document

## PID Sign-Off

|  |  |
| --- | --- |
| **Student No.** | **30000573** |
| **Student Name** | **Sebestyen Morris** |
| **Email** | [**Js000573@student.reading.ac.uk**](http://Js000573@student.reading.ac.uk) |
| **Degree programme** (BSc CS/BSc CSwIY) | **BSc Computer Science** |
|  |  |
| **Supervisor Name**  *(Consultation with supervisor is mandatory)* | Dr. Luis Patino |
|  | Supervisor to sign PID form on Bb (grade centre) |
| **Date** | **28th September 2023** |

# SECTION 1 – General Information

## Project Identification

|  |  |
| --- | --- |
| **1.1** | **Project Title** |
|  | The Autonomous R.A.T. Machine |
| **1.2** | **Please describe the project with key-phrases (max 5)** |
|  | LLM powered automatic remote access trojan for penetration testers. |
| **1.3** | **E-logbook maintenance agreed with supervisor**  *Use Google doc, OneDrive, or any mobile App whereby you will be able to generate a PDF copy* |
|  | Word document in onedrive. |
| **1.4** | **GitLab link for maintain source code and research data**  *Any change in GitLab link and Source code repository MUST be explicitly mention in final report* |
|  | <https://csgitlab.reading.ac.uk/js000573/final-year-project> |

# SECTION 2 – Project Description

|  |  |
| --- | --- |
| **2.1** | **Summarise the project’s background in terms of research field /application domain (max 100 words).** |
|  | The project researches the applications of artificial intelligence, specifically large language models to aid and speed up the work of penetration testers. The goal is to make a self contained device that will use its own reasoning capabilities in order to gain an initial foothold on a network for a penetration tester. The main area of research for this project understanding how to guide a GPT based agent to complete such a complex task. |
| **2.2** | **Summarise the project aims, objectives and outputs (max 250 words).** These aims, objectives, and outputs should appear as the tasks, milestones and deliverables in your project plan (fill out Section 3). |
|  | The aim of this project is to understand to what extent a simple language model can be used to carry out complex tasks. The project will compose of a task manager, which guides the model into carrying out a structured plan and numerous language model based agents which can accomplish tasks in whichever way is most fit for the situation. If successful, the project should be able to conduct a penetration test on a wide variety of systems without explicitly being programmed to do so.   * Understand how to make a reliable GPT based agent using langchain (or similar alternative libraries) * Research whether the performance of the models can be improved via methods such as fine-tuning or adding vector databases of security information. * Implement a task manager system to keep track of the process of the penetration test and guides the model to its next steps. * Implement the software side of the program. * Implement a reporting mechanism that logs all actions taken by the device and sends it to the operator. * Create a raspberry pi based self contained device to house the system. * Create a way for the system to communicate with the operator remotely. This will most likely require an LTE connection and a web interface or app to control the device from. * Evaluate the performance by testing it on simulated networks and systems. * Document the entire process to aid with report creation. |

|  |  |
| --- | --- |
| **2.3** | **Initial project specification – roughly indicate key features and functions of your finished program/application. Indicate possible method, data source, technology etc. (max 400 words)** (Sensible and relevant Charts, Table, and Figures can be used) |
|  | The finished project should take the shape of a device based on a raspberry pi acting as an autonomous ‘Remote Access Trojan’[[1]](#footnote-2). The device will use an LTE connection to communicate with an operator. If the device gets connected to any network, with the operators permission it should start trying to gain control of as many devices on that network as it can. Once successful, it should set up a stable entry-point to the system and relinquish control to the operator.  Here are some of the key features needed to achieve this:   1. Basic information gathering: Before starting to work, the system needs to be aware of what type of network its on and how to begin scanning for other devices. 2. Network scanning: The system should be capable of doing an intensive scan of the entire network its connected to. It should be able to identify all devices, what they might be used for, what network ports are open, and whether they would make a good target to exploit. 3. Target switching: The project should be able to have multiple target devices that it can try to gain control over. When a regular language model is given multiple tasks, it tends to choose one and only do that. This needs to be overcome by a task managing system. 4. Exploitation: The project needs to have at least a basic ability to exploit systems and gain control of them. It is unclear at the moment how affective it will be at exploiting devices, but it should be able to succeed on easier, more automated tasks. 5. Setting up control for the operator: If the exploitation is successful, the system needs to somehow set up a back-door for the operator to use the exploited device. 6. Operator control: The operator should have a live view of what the system is doing and a reliable way to stop it or guide it in a different direction if need be. To make sure that issues on the target network don’t interrupt this, it will be done with an LTE adapter to the raspberry pi. 7. Logging and reporting: Everything this system does needs to be reported, both in real-time for the operator and in some form for it to be included in a penetration testing report. |
| **2.4** | **Describe the social, legal and ethical issues that apply to your project. Does your project require ethical approval? (If your project requires a questionnaire/interview for conducting research and/or collecting data, you will need to apply for an ethical approval)** |
|  |  |
| **2.5** | **Identify the items you may need to purchase for your project. A cost upto £200 can be applied (include VAT and shipping if known). You need to have consent of your supervisor. Your request will be assessed by the department.** |
|  | USB – LTE adapter for communication: 20£  M.2 LTE module: 40£  OpenAI API credits: 140£ |
| **2.6** | **State whether you need access to specific resources within the department or the University e.g. special devices and workshop** |
|  | no |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **SECTION 3 - Project Plan** | | | | | | | | | | | | | |
| Please provide your project plan.  Below is an example project plan, you can use any tool or software to generate yours. | | | | | | | | | | | | | |
| **Project stage** [this is only indicative – write your own stages] | **START DATE: ../../…. <enter the project start date here>****Project Weeks** | | | | | | | | | | | | |
| 0-3 | 3-6 | 6-9 | 9-12 | 12-15 | 15-18 | 18-21 | 21-24 | 24-27 | 27-30 | 30-33 | 33-36 | 36-39 |
| 1 Background Research |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1.1 Research best ways to create autonomous agents from LLMs |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1.2 Research what tools can be used by agents or if some custom ones have to be made |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1.3 Can fine tuning and vector databases help make it more accurate |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1.4 Develop proof-of-concept prototype tools or agents to check if concepts are feasable |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 Analysis/Design |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2.1 Design and draw out the overall layout of the system |  |  |  |  |  |  |  |  |  |  |  |  |  |
| .. |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3 Develop prototype |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Develop the task manager system |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Implement the scanning and reconnaissance phase |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Implement exploitation stage |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Create remote control interface for operator |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Put final device together |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 Testing/evaluation/validation |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Testing and evaluating performance |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 Assessments |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Report writing |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |

1. A computer security term for software or a device that gives the operator unauthorized, remote control over a system or network. These programs are commonly used by penetration testers to prove the impact of their work and to help them conduct further tests on internal devices of the network. [↑](#footnote-ref-2)