UNIVERSITY OF
WEST LONDON

RESEARCH ETHICS AND INTEGRITY

RISK ASSESSMENT FORM

FORM A

FOR UNDERGRADUATE AND TAUGHT POSTGRADUATE STUDENTS

FOR OFFICE USE ONLY

To be completed by module leader/supervisor [subject to confirmation by SCREP]

In the opinion of the module leader/supervisor this application falls into

CATEGORY I [ ]

CATEGORY 2 [ ]

Please fill in this form, then SAVE IT AS A PDF and submit as instructed by your supervisor with your project proposal

Your name: (first) ARAHARAN (last) LOGANAYAGAM

Student number: 21524785 Your email address: 21524785@student.uwl.ac.uk

Name of supervisor: Dr.Waqar Asif

Title of project:

Blockchain-and ML based Malware Detection and Integrity Checking: A Decentralized Approach

Date:30/05/2023

SECTION A

PROJECT DESCRIPTON

Please answer the following questions:

1. **Do you intend to involve human participants in the conduct of your research? If no, please skip questions 1a & 1b.**

   ☐ Yes ☒No

   1a. Does your research involve vulnerable adults (who are or may be for any reason unable to take care of themselves, or unable to protect themselves against significant harm or exploitation) or under-18s?

   ☐ Yes ☐No

   1b. Could your research potentially expose you, anyone assisting you, or participants to physical, psychological and/or emotional harm? (see Section B, Question 9)

   ☐ Yes ☐No

2. **Will your research involve travelling to geo-politically unstable regions/countries (e.g. areas affected by war, civil unrest, natural disasters, or listed as unadvisable to travel by the UK government)?**

   ☐ Yes ☒No

3. **Will your research involve access to security-sensitive material? (see the University's Research Ethics Code of Practice 2018 for a definition of security-sensitive materials and Section B, Question 9 of this form)**

   ☐ Yes ☒No

This proposal ***must be completed with the assistance of your supervisor*/module leader/tutor**. You can change the size of the boxes (below) by typing or deleting as necessary.

It is very important to convey **with clarity**:

- Your research questions/the problem/the theme or topic you are investigating (what you are proposing to do and to find out or to create)

- The methodology or technical approach (for projects comprising in whole or in part the creation of an artefact) you will adopt – methods, number of participants, who the participants (if any) will be, survey instruments used, technology and equipment employed etc.; and what questions you are planning to ask your respondents (if applicable); how you will deal with technical challenges.

SECTION B

Only complete if you answered YES to Q1 in Section A.

	WHERE APPROPRIATE TO YOUR CHOSEN TOPIC/RESEARCH:		YES	NO	N/A
1	Will you describe in writing the main procedures to participants in advance, so that they are informed about what to expect? A copy of this must be attached to this application				
2	Will you tell participants that their participation is voluntary?				
3	Will you obtain written consent for participation and include within this that they have a right to withdraw at any point? A copy of this must be attached to this application				
4	If the research is observational, will you ask participants for their consent to being observed?				
5	With questionnaires, will you give participants the option of omitting questions they do not want to answer?				
6	Will you tell participants that their data will be treated with full confidentiality and that, if published, it will not be identifiable as theirs? This should be evidenced in the consent form and (if applicable) with a signed copy of UWL's data management form, attached to this application.				
7	Will you debrief participants at the end of their participation (i.e. give them a brief explanation of the study)? A copy of this must be attached to this application				
8	Will your project involve deliberately misleading participants in any way?				
9	If you answered YES to Question 1b (section A) give details on a separate sheet and state what you will tell your participants to do if they should experience any problems (e.g. who they can contact for help).				
10	Do participants fall into any of the following vulnerable groups? If they do, please and **tick box 2** overleaf.	Schoolchildren (under 18 years of age)			
		People with learning or communication difficulties			
		Patients			
	Note that you may also need to obtain satisfactory DBS clearance (or equivalent for overseas students).	People in custody			
		People engaged in illegal activities (e.g. drug-taking)			

		Any other groups who could be reasonably argued as representing any form of vulnerability – please specify			

SECTION C

	WHERE APPROPRIATE TO YOUR CHOSEN TOPIC/RESEARCH:	YES	NO	N/A
11	Will you be accessing materials which may be considered security-sensitive under the Counter Terrorism Act (2015)?		✗	
12	Does your project involve work with animals? If yes, please **tick box 2** below.		✗	

[Note: N/A = not applicable]

There is an obligation on the researcher to bring to the attention of the School Ethics Panel any issues with ethical implications not clearly covered by the above checklist.

PLEASE TICK **EITHER** BOX 1 OR BOX 2 BELOW AND **PROVIDE THE DETAILS REQUIRED** IN SUPPORT OF YOUR APPLICATION. THEN SIGN THE FORM.

Please tick

1.	I consider that this project has **no** significant ethical implications to be brought before the School Ethics Panel.	✓

2.	I consider that this project **may** have ethical implications that should be brought before the School Ethics Panel, and/or it will be carried out with children or other vulnerable populations.	

I have received guidance on ethical research practices relevant to my subject as part of my preparation for this module.

Signed ______Araharan Loganayagam_____ Print Name ______Araharan Loganayagam____________

Date __30/05/2023______

(UG Researcher(s))

Signed ______________________________ Print Name _________________________

Date ________________

(Supervisor)

PROJECT OUTLINE

Your name: (first) …ARAHARAN…… (last) …LOGANAYAGAM……

Student number: …21524785…… Your email address: 21524785@student.uwl.ac.uk

Name of supervisor _______Dr. Waqar Asif______________________

Title of project __ Blockchain-and ML based Malware Detection and Integrity Checking: A Decentralized Approach _____

Date:__30/05/2023______

Introduction to the research

Background to research topic area (with references where applicable). 200 words approx. Include **Aims and Hypothesis, Research Question(s) of the dissertation/project or an outline of the aims and the context of the creative artefact (where the project is a creative artefact)**

Background

In recent years, malware attacks have become increasingly prevalent, sophisticated and pose a significant threat to the security of computer systems and networks. To address this challenge, researchers have developed various malware detection and integrity checking techniques, ranging from traditional signature-based methods to more advanced machine learning and blockchain-based approaches. While these techniques have been successful in detecting and preventing malware attacks, the ever-evolving nature of malware requires continued research and innovation to stay ahead of threats. This literature review will focus on the current state of malware detection and integrity checking techniques, specifically exploring the potential of a decentralized approach based on blockchain and machine learning. By decentralizing malware detection and integrity checking, blockchain technology can provide a transparent, secure, and tamper-proof environment for detecting and preventing malware attacks. Machine learning can also enhance the accuracy and effectiveness of malware detection by analysing large amounts of data and identifying patterns that may not be visible to traditional signature-based methods. The traditional signature-based methods, anomaly-based detection, and machine learning-based methods. It will then explore how blockchain and machine learning can be used together to create a decentralized approach for malware detection and integrity checking. Finally, the review will analyse the potential benefits of a decentralized approach, including increased transparency, security, and efficiency.

Aim

1. Develop a robust ML model: Build and train an ML model capable of accurately identifying and classifying malware based on patterns, behaviors, or other relevant characteristics. The model should be adaptable to new malware threats and capable of continuous learning.
2. Implement a decentralized architecture: Design and implement a decentralized architecture where the workload of malware identification and integrity checking is distributed across multiple nodes or participants in the network. Ensure scalability, fault tolerance, and efficient processing of the ML algorithms.
3. Integrate blockchain technology: Incorporate blockchain technology into the system to leverage its decentralized, transparent, and immutable properties. Utilize the blockchain to securely store and validate the results of malware identification and maintain the integrity of the training data.
4. Explore consensus mechanisms: Investigate different consensus mechanisms used in blockchain networks (e.g., Proof of Work, Proof of Stake) and determine the most suitable mechanism for achieving distributed consensus in the context of malware identification and integrity checking. Assess the impact of consensus mechanisms on the accuracy and reliability of the system.
5. Evaluate system performance: Conduct comprehensive performance evaluations to assess the effectiveness, efficiency, and reliability of the proposed system. Measure the accuracy of malware identification, the speed of processing, the scalability of the decentralized architecture, and the robustness of data integrity measures.
6. Validate real-world applicability: Validate the research findings by testing the system with real-world malware samples and scenarios. Analyze the system's performance in detecting and mitigating real-world threats and compare it with existing centralized malware identification solutions.

Hypothesis

1. Machine Learning (ML) for Malware Identification: The hypothesis assumes the usage of ML algorithms to identify and classify malware based on patterns, behaviors, or other characteristics. ML models can be trained using labeled datasets and can continuously learn and adapt to new malware threats.
2. Decentralization: By employing a decentralized architecture, the hypothesis suggests distributing the workload of malware identification and integrity checking across multiple nodes or participants in the network. This decentralization can enable faster processing, scalability, and fault tolerance, as well as reduce the reliance on a single point of failure.
3. Blockchain Technology: The hypothesis proposes integrating blockchain technology to enhance the security and trustworthiness of the ML-based malware identification system. Blockchain provides a decentralized, transparent, and tamper-resistant ledger where each transaction or operation is recorded in a block and linked together in a chain. This immutability ensures the integrity and transparency of the system.
4. Distributed Consensus: Blockchain networks typically employ consensus mechanisms (e.g., Proof of Work, Proof of Stake) to validate and agree upon the order and content of transactions. By leveraging distributed consensus, the hypothesis assumes that the accuracy and reliability of the malware identification process can be improved, as multiple nodes in the network validate the results and agree on the correctness of identified malware.
5. Data Integrity: In the proposed system, the hypothesis suggests that the immutable nature of blockchain can be leveraged to ensure the integrity of the data used for training ML models and verifying the accuracy of malware identification results. Once data is recorded on the blockchain, it becomes practically impossible to modify or tamper with, which helps establish trust in the system.

Questions in traditional approach

Traditional malware detection systems rely on signature-based detection methods and pattern matching techniques to identify known malware threats. However, these methods have limitations that make them less effective in detecting new and unknown types of malwares, which are increasingly common in the real world. • they can only detect known threats that have already been identified and added to a signature database. This means that if a new or modified malware variant is released, it may go undetected by traditional detection methods until it is added to the signature database. • As well as There are several challenges in making generalized malware detection models. Malware is constantly evolving, with new variants and attack methods emerging regularly. This makes it challenging to build a generalized model that can detect all types of malwares. • Adversarial attacks are techniques that are used to bypass or manipulate machine learning models. In the case of malware detection models, adversaries may use these attacks to evade detection. • Malware detection models need to be fast and efficient, particularly when used in real-time environments. The computational overhead associated with some machine learning algorithms can make them impractical for use in some environments.

Method

Outline all methodological issues **for any projects requiring human participants and data collection** - e.g. Design, participants, questionnaires, tests, method of data collection. *Who are you working with? How? What Measures? What interventions/manipulations? What controls?*

1. Training Data for Machine Learning:

- Malware Samples: A diverse dataset of malware samples is necessary to train the machine learning model. This dataset should include a variety of malware types, families, and variants. Obtaining a representative and comprehensive collection of malware samples may involve collaborating with cybersecurity organizations, research institutions, or leveraging publicly available datasets.
- Labeled Data: Each malware sample in the training dataset should be labeled with the corresponding malware type or family. This labeled data will serve as the ground truth for training the ML model to classify and identify malware accurately. The labeling can be done manually by experts or by utilizing existing labeled datasets.

If your research entails exclusively the **consultation of published documents**, books, articles or other work in the public domain please state this under the heading ‘materials’.

In the case of **creative artefacts** (such as audio-visual, audio or visual outputs or production such as a film, video, audio recording, composition, screenplay, piece of creative writing, performance, exhibition, screening, photograph, body of photographic work, painting, sculpture, installation, design or software) please use relevant sections of your dissertation/project proposal to place in the section below to which they most closely pertain.

Students and supervisors may also find it helpful to cross-refer to the Health and Safety clearance documents for any level 6 research projects which need to be completed by students ***in certain fields***. Sections completed by students for the latter forms may be suitable to be repeated below.

Research design or schedule (for creative artefacts)

Month 1:

Week 1:

Familiarize yourself with existing research on blockchain-based ML systems for malware identification and integrity checking.

Define the specific objectives, scope, and requirements of your project.

Week 2:

Gather and curate a diverse dataset of malware samples for training the ML model.

Preprocess and prepare the malware dataset by extracting relevant features and ensuring data quality.

Week 3:

Develop and train the machine learning model using the prepared dataset.

Evaluate the performance of the model using appropriate metrics and techniques.

Week 4:

Design the decentralized architecture for the malware identification and integrity checking system.

Determine the necessary blockchain components and consensus mechanism to be implemented.

Month 2:

Week 1:

Implement the decentralized architecture and integrate the trained ML model into the system.

Set up the blockchain network and configure the necessary components.

Week 2:

Develop the smart contracts or protocols required for recording the malware identification results and other relevant data on the blockchain.

Implement the consensus mechanism chosen for achieving distributed consensus.

Week 3:

Test and validate the functionality of the integrated system.

Conduct initial performance evaluations and address any issues or bugs.

Week 4:

Improve the system's scalability, fault tolerance, and efficiency by optimizing the decentralized architecture and the blockchain implementation.

Prepare for the data collection phase by setting up necessary data storage and retrieval mechanisms.

Month 3:

Week 1-3:

Begin the data collection phase by processing and analyzing real-world malware samples using the developed system.

Record the transaction data on the blockchain, including malware identification results, timestamps, and relevant metadata.

Week 4:

Evaluate the performance and accuracy of the system in detecting and identifying malware using the collected data.

Analyze the results and compare them with existing centralized malware identification solutions.

Month 4:

Week 1:

Analyze the performance and effectiveness of the consensus mechanism in achieving distributed consensus for malware identification and integrity checking.

Evaluate the system's scalability and fault tolerance under different workloads.

Week 2-3:

Write the final research report, documenting the methodology, findings, and conclusions of the project.

Include performance evaluations, analysis of results, limitations, and recommendations for future work.

Week 4:

Prepare and deliver a presentation summarizing the project's objectives, methodology, and key findings.

Review and finalize the research report, ensuring it meets the required standards.

Participants, including (where applicable) collaborators in the making of creative artefacts

Materials (to include locations and objects/resources)

Procedure or details of technical aspects of creative production

Analysis

Please complete this section **only if your project requires written analysis** to be submitted as the assessment or as part of it. If the project you are undertaking comprises a creative artefact such as a film or body of photographic work please type 'Not applicable' in this box

CLEARLY describe the method of analysis you are going to use. Is it *qualitative* or *quantitative?*

The method of analysis for this project can incorporate both quantitative and qualitative approaches, depending on the specific aspects being examined.

Quantitative Analysis:

- Performance Metrics: Quantitative analysis can involve measuring performance metrics such as accuracy, precision, recall, F1-score, and detection rates to evaluate the effectiveness of the machine learning model in identifying malware.

- Throughput and Scalability: Quantitative analysis can assess the throughput and scalability of the decentralized architecture by measuring the number of transactions processed per second, latency, and resource utilization under varying workloads.

Qualitative Analysis:

- User Experience and Feedback: Qualitative analysis can involve gathering user feedback through surveys, interviews, or user testing sessions to assess the usability, user experience, and user satisfaction with the system.

- System Robustness: Qualitative analysis can involve evaluating the system's robustness by subjecting it to real-world malware samples and assessing its ability to accurately identify and handle various types of malwares.

For students completing a **Dissertation** you should be able to refer to what you have learned in the Research Methods component of your study.