## Initial Project Overview

## SOC10101 Honours Project (40 Credits)

### Title of Project:

**Machine Learning-Driven Intrusion Detection System for SME Networks**

### Overview of Project Content and Milestones

This project aims to design, implement and evaluate a lightweight intrusion detection system (IDS) that applies machine learning techniques to detect anomalous traffic within small or medium enterprise (SME) networks – in particular HTTPS traffic and DDoS attacks.

The research will focus on comparing and evaluating existing IDS models and approaches, identify their limitations in SME contexts, and develop a prototype tailored for HTTPS anomaly detection for smaller scale environments.

The project will progress through.

* Oct. - Nov.: A literature review on HTTPS-focused IDS and gaps analysis; Draft the literature review chapter
* Nov: Collection and preprocessing of HTTPS and DDoS traffic datasets
* Dec. - Jan.: Design and implementation of the ML-based IDS; Draft the methodology chapter
* Feb.: Empirically evaluate ML-based IDS with pre-collected data
* Mar.: Draft the remaining chapters of the dissertation
* April: Taking feedback and finalising the dissertation

### Motivation and Research Context:

Small and medium enterprise (SME) networks face growing cyberattack risks but lack affordable intrusion detection solutions, particularly in anonymous HTTPS traffic, which dominates web-based services used by small businesses. As a result, there is growing need for adaptable, low-resource IDSs capable of detecting and classifying abnormal network behaviour using accessible, open-source methods.

This project is motivated by this emerging security gap and seeks to explore how AI and machine learning can be leveraged to improve detection accuracy and efficiency in small scale environments. The work combines applied experimentation with a research driven evaluation of current IDS models and techniques, contributing to an unexplored area of cybersecurity research focused on lightweight, intelligent defences for constrained SME contexts.

### The Main Deliverable(s):

This project will produce a set of deliverables designed to address the cybersecurity needs of small and medium enterprise (SME) networks, with a focus on developing a practical, lightweight intrusion detection system (IDS) tailored for small businesses.

This project will produce deliverables addressing the cybersecurity needs of small and medium enterprises (SMEs), focusing on a lightweight, machine learning-based IDS for HTTPS/Web traffic. The primary deliverable is a functional prototype that ingests HTTPS datasets (e.g., CIC-IDS2017 TLS flows) and applies models like Random Forest or Isolation Forest to detect anomalies such as data exfiltration, injection attacks, DDoS, and brute force attempts, using open-source tools for accessibility. A comprehensive dissertation will document the research context, literature review, methodology, implementation, and evaluation.

An organized repository will include:

* HTTPS-focused datasets (e.g., CIC-IDS2017 TLS flows)
* Test scenarios simulating real world SME network attacks (e.g., DDoS, brute force, data exfiltration).
* Performance metrics (e.g., accuracy, false positives/negatives rates)
* Comparative results between ML models

These deliverables aim to balance academic rigor with real world utility, supporting the development of an affordable, deployable IDS solution.

### The Target Audience for the Deliverable(s):

The deliverables are designed to serve a diverse audience invested in securing SME networks, particularly small businesses with limited resources. The primary audience includes:

* SME administrators, who need lightweight IDS solutions to protect web-based services.
* Independent security consultants and freelancers, who work with small-scale / SME clients and require affordable solutions.
* Academic researchers, who are exploring scalable applications of ML in IDS development.

By focusing on web-based applications like HTTPS traffic, which dominates modern small business operations, the project ensures relevance to practical and academic stakeholders while remaining accessible to non-technical users.

### The Work to be Undertaken:

The project adopts a research-driven yet practical methodology to design, train, and evaluate a machine learning-based IDS tailored for SME environments, with an emphasis on detecting anomalies in HTTPS traffic and DDoS attacks. The work will begin with a comprehensive literature review of current IDS methodologies, focusing on machine learning approaches (e.g., Random Forest, Isolation Forest) and their applicability to HTTPS traffic and DDoS/brute force scenarios in small-scale networks, to identify limitations and research gaps. Publicly available datasets, such as CIC-IDS2017 (with TLS flows) and UNSW-NB15, will be collected and pre-processed to ensure data consistency and quality for HTTPS-focused anomaly detection and DDoS/brute force scenarios, with potential exploration of SMTP datasets for phishing detection if feasible. The core implementation involves training and evaluating ML models to detect suspicious activities, analysing metrics such as:

* Detection accuracy,
* False positive and negative rates,
* Computational efficiency for SME deployment.

Results will be synthesized to reflect on findings in relation to existing IDS research, identifying improvements or future research directions. This structured approach ensures the project delivers a practical prototype while contributing to the academic understanding of lightweight IDS solutions for small businesses.

### Additional Information / Knowledge Required:

To successfully execute this project, a range of research and technical skills will be needed. Including:

* An understanding of machine learning techniques, such as Random Forest, applied to cybersecurity for anomaly detection.
* Familiarity with HTTPS protocol analysis and TLS traffic inspection without decryption, with potential exploration of SMTP for phishing detection.
* Proficiency in network traffic analysis, Linux networking, and packet inspection tools like Wireshark or Zeek to process and analyse datasets.
* Knowledge in data preprocessing and model evaluation to ensure robust ML performance.
* Awareness of current research trends in ML based IDS, particularly for HTTPS traffic in resource constrained SMEs.

### Information Sources that Provide a Context for the Project:

The project will draw on a variety of credible sources to provide context and support its development. These include:

* Academic and industry research:
  + Recent publications on machine learning-based IDS, particularly those addressing web-based traffic analysis in resource constrained SMEs.
* Publicly available datasets:
  + CIC-IDS2017 (Canadian Institute for Cybersecurity), which includes TLS flows for HTTPS anomaly detection
  + UNSW-NB15 (Australian Centre for Cyber Security), containing HTTPS-related attack scenarios.
* Open-source IDS tools
  + Like Zeek, Snort and Suricata, which provide context for web-based traffic analysis.

These sources will ground the project in current research and ensure the prototype is informed by real world data and tools, enhancing its relevance for SME networks.

### The Importance of the Project:

This project addresses a critical cybersecurity gap by developing a lightweight, ML-based IDS tailored to resource constrained SME networks, which are increasingly targeted by cyberattacks but lack enterprise-grade solutions. By focusing on HTTPS traffic, which underpins web-based services used by small businesses, aswell as a secondary focus on DDoS / brute force attacks, the project aligns with industry demand for accessible, open-source security tools.

It contributes to academic research by evaluating ML models in the context of resource constrained environments, advancing the understanding of scalable and efficient IDS solutions.

By combining applied experimentation with research driven insights, the project offers both immediate utility for SMEs and a foundation for future innovations in lightweight IDS development.

### The Key Challenge(s) to be Overcome:

Developing a lightweight, machine learning-based IDS for SME applications, presents several challenges:

* Data Quality and Preprocessing:
  + Cleaning and formatting datasets (e.g., CIC-IDS2017 TLS flows) to extract metadata like packet size.
* Scope Management:
  + Prioritizing HTTPS anomaly detection, with DDoS and brute force as secondary focuses and SMTP/malware as optional, to avoid scope creep within the project timeline.
* Computational Constraints:
  + Ensuring ML models are efficient for deployment on low cost SME hardware
* Research Integration:
  + Aligning the HTTPS and DDoS focused prototype with literature review findings to support project goals.

By addressing these challenges, the project will deliver a practical IDS prototype that meets the cybersecurity needs of a small business while contributing to academic research.