COMP4336/9336 Mobile Data Networking: 2024 Term 3
Individual Term Project/Assignment
Weighting: 25%
Due: 5pm Fri 08 November 2024 (Week 9)

Project Conditions

- This is an **individual** assignment. The work you submit must be entirely your own work. Submission of work even partly written by any other person or AI is not permitted.
- Do not provide or show your assignment work to any other person, other than the teaching staff of COMP4336/9336.
- Do not publish your assignment solution (data/code) via the Internet. For example, do not place your assignment in a public GitHub repository. Sharing, publishing, or distributing your assignment work even after the completion of COMP4336/9336 is **not** permitted.

Violation of any of the above conditions may result in an academic integrity investigation, with possible penalties up to and including a mark of ZERO in COMP4336/9336, and exclusion from future studies at UNSW.

Project Topic: Data-Driven WiFi Enhancement

Objective

This project is designed to deepen your understanding of the factors that influence retransmissions in WiFi networks, with a focus on how these retransmissions impact network performance. Through this project, you will gain hands-on experience in data collection, performance analysis, and applying basic machine learning techniques to real-world network data.

Learning Outcomes:

- **Practical Data Collection:** You will learn how to effectively capture and analyse real-world WiFi traffic data, identifying key patterns related to retransmissions.
- **Performance Insight:** By analysing the impact of retransmissions on network performance, you will develop a deeper understanding of the challenges in WiFi network management.
- **Machine Learning Application:** You will apply machine learning to predict retransmission events, gaining insight into the predictive power of various network factors.
- **Critical Thinking and Innovation:** The project will culminate in designing protocol enhancements that address the identified issues, fostering your ability to innovate within existing network frameworks.
- **Real-World Application:** The skills and knowledge gained from this project will be directly applicable to solving real-world networking challenges, preparing you for careers in network engineering, cybersecurity, or research and development.

Wireless Data Collection
Data Processing & Performance Analysis
Machine Learning Evaluation
Protocol Enhancement Design

Fig 1. Project Tasks

Project Tasks and Components

1. **Data Collection:**

   You will use Wireshark or other protocol analysers to collect WiFi traffic data from operational WiFi networks. Your goal will be to gather a rich dataset that includes both normal and retransmitted packets. Pay attention to the following:

   - **Time and Location Diversity:** You should capture data from at least **two** crowded locations where you can observe significant WiFi retransmissions. Note that in light-traffic scenarios, there would be not many WiFi collisions and retransmissions, making data collection not particularly useful for your project. Therefore, it is crucial to choose locations with high network activity and high density of WiFi Access Points (APs), such as libraries, UNSW CSE buildings, or similar busy areas. For a given location, collect data across different times of the day and over multiple days spanning different weeks. You have the flexibility to choose the locations and times for your data collection.
   - **Frequency and Channel Coverage:** Collect data in **2.4 GHz WiFi**, covering as many channels as possible within 2.4 GHz bands. Collect data from as many APs as possible. If you find that your Windows laptop does not support capturing WiFi from other APs, you may have to use a MAC device (laptop) or a Raspberry PI (Model 3/4/5).
   - **Data Volume**: Collect enough data to ensure that your conclusions are statistically valid and robust enough for training machine learning models. Adequate data will help the model generalize well and minimize the risk of overfitting.

2. **Performance Analysis:**
   Once the data is collected, you will analyze the impact of WiFi collisions and retransmissions on network performance. This involves identifying retransmitted packets and calculating the delay caused by these retransmissions. You will explore how retransmission frequency correlates with overall network latency, throughput, and efficiency, and visualize these relationships to uncover patterns and insights.
   Task details:
   - **Data Processing:** You may need to process the data before performing any analysis. Data processing includes merge all the data files you have collected, filtering unwanted data, exporting the data to file and ready to use later. Remove all the non-related packets by using filter and identify all the retransmitted packets. Hint: In Wireshark, you can filter 802.11 packet by using filter: `wlan` and find the retransmitted packets with filter: `wlan.fc.retry == 1`

   - **Retransmission Analysis:**
     To analyze the retransmissions in your dataset, use Wireshark or an equivalent packet sniffing tool to identify and filter retransmitted packets. Focus on the 802.11 frame retransmission indicators. For each packet, track how many times it has been retransmitted (e.g., once, twice, etc.). Present the retransmission distribution in your dataset, showing how often packets are retransmitted. You can display this using a bar chart or histogram,

where the x-axis represents the number of retransmissions and the y-axis represents the number of occurrences.

- **Network Performance Analysis:**
  Do the network performance analysis by choosing one of the most congested networks in your dataset. Next, calculate key network performance metrics such as latency, throughput, and efficiency.
  - **Latency**: Locate a data packet and the corresponding ACK frame. Calculate the latency as the difference between the ACK packet's timestamp and the data packet's timestamp.
  - **Throughput**: Calculate the throughput as the total amount of data successfully transmitted over a given period. You can compute this by summing up the **data packet** sizes filtered for a given device (MAC address) and dividing by the duration of the capture.
  - **Efficiency**: Evaluate the efficiency by comparing the amount of useful data transmitted to the total number of transmissions, factoring in retransmissions.

  To present your findings, create graphs and charts that show the relationship between retransmission frequency and network performance metrics such as latency, throughput, and efficiency. Use visualization tools like Matplotlib or Seaborn to create plots that clearly depict these relationships.

3. **Machine Learning Evaluation:**
   You will apply machine learning techniques to evaluate the factors that influence retransmissions. By extracting features such as signal strength, packet size, and channel utilization, you will train a machine learning model to predict the likelihood of packet retransmission. This model will help you understand which factors are most critical in determining whether a packet is successfully transmitted or needs to be retransmitted.
   - **Feature Extraction:** Extract relevant features from your dataset, such as signal strength, packet size, and channel utilization, to serve as inputs for your machine learning model.
   - **Model Training and Evaluation:** Train some basic machine learning models e.g., kNN, Random Forest, and SVM, to predict packet retransmissions. Evaluate the models' performance using a portion of your dataset and analyze the results.
   - **Model Interpretation:** Discuss which factors/features are most influential in predicting retransmissions and the significance of these factors for network performance.

Resources for machine learning:

- Books & Online Documents
  - Scikit-learn Documentation: The official documentation is comprehensive, providing guides, examples, and references that are invaluable for both beginners and advanced users.
  - Python Machine Learning by Sebastian Raschka
- Online Courses
  - Scikit-learn in Python for Machine Learning
  - Scikit-learn for Beginners
  - Coursera: Applied Machine Learning in Python: offered by the University of Michigan

4. **Protocol Enhancement Design:**
    - Based on the insights gained from your machine learning analysis, you will propose **enhancements** to the WiFi protocol aimed at reducing retransmissions and improving overall network performance. While implementing your proposed enhancement is not required, you should demonstrate your proposal's feasibility through analysis and discussions. Here are some example enhancements you might consider (but don't feel limited to these):
        - **Adaptive Retransmission Mechanisms:** Design mechanisms that adjust retransmission strategies based on real-time network conditions, such as varying the backoff time or retransmission count based on signal strength or congestion levels.
        - **Dynamic Channel Allocation:** Propose a method for the AP to dynamically switch the channel that reduces the likelihood of retransmissions.
        - **Error Control Enhancements:** Suggest improvements to error detection and correction mechanisms to prevent the need for retransmissions under certain conditions.
    - **Critical Thinking:** Explain the rationale behind your proposed enhancements, considering trade-offs such as complexity, scalability, and potential impacts on existing WiFi standards. Evaluate how these changes might be implemented in real-world networks and what challenges might arise.

Submission Requirements

1. **Collected Dataset:** Submit the dataset in a zip file, organized according to the template to be released. If the file size exceeds Moodle's upload limit, upload it to Google Drive or GitHub and provide the link in your submission (make sure you give access permission to your tutor). No modifications can and should be made to the dataset after the submission deadline.
2. **PDF Report:** Submit a single PDF report (maximum 10 pages) that includes:
    - A detailed explanation of the data collection process and detailed metadata of your datasets.
    - Analysis of the impact of retransmissions on network performance, including visualizations and insights derived from the analysis.
    - A discussion of the machine learning evaluation, covering feature extraction, model training, and model performance comparison.
    - The design and critical analysis of proposed protocol enhancements, including the rationale behind your decisions, trade-offs considered, and feasibility for real-world implementation.
3. **Jupyter Notebook and Code:**
   Submit all code used for data analysis and performance evaluation, either as standalone scripts or within the Jupyter notebook. Include a Jupyter notebook (Python 3) that demonstrates the machine learning process, from importing the dataset to training and evaluating models using scikit-learn (or other machine learning libraries). Ensure that the notebook is well-documented with explanations of each step.

Submission Process:

Upload the PDF report, code files, and Jupyter notebook via the designated Moodle submission link by 5 pm, Fri 08 November 2024. Late submissions will incur a penalty of 5% per day. No submissions will be accepted after 15 November.

Marking Rubic:

Task	Excellent (85-100%)	Average (50-84%)	Poor (0-49%)
Data Collection (25%)	Rich dataset submitted in the requested format, covering at least 2 different locations, across wide range of times(day and night, at least 3 different days) for each location, with data spanning different hours, days and multiple weeks.	Dataset meets minimum requirements but lacks diversity in locations, times, or frequency coverage. Format may be inconsistent or lacks adherence to provided template.	Incomplete dataset with insufficient locations, times, or frequency coverage. Format is incorrect or dataset is poorly organized, hindering further analysis.
Data Analysis (25%)	Professional analysis with detailed procedures. Outcomes are clear, easy to understand, and well-presented with insightful visualizations. Analysis connects well with course content.	Analysis is adequate but lacks depth or clarity. Some visualizations may be missing or unclear. The connection to course content is present but not strongly articulated.	Analysis is superficial or incorrect. Visualizations are unclear or missing. The report is difficult to follow and lacks connection to course content.
Machine Learning (25%)	Dataset is preprocessed correctly. At least two machine learning algorithms are compared. Perform analysis to prove has adequate data collection. Clear examples of model usage and thorough analysis of performance are provided.	Basic preprocessing is done, but comparison of models is limited or not well-explained. Some examples are unclear or missing, with limited analysis of model performance.	Preprocessing is incorrect or missing. Only one model is used without comparison, or analysis is severely lacking or incorrect.
Protocol Enhancement Design and Discusion (25%)	Demonstrates strong understanding of wireless protocols and proposes innovative enhancements with clear feasibility and benefit analysis.	Some understanding of wireless protocols is demonstrated, proposed enhancements appear to have some benefit, but lacks comprehensive discussion.	Limited or no understanding of wireless protocols is demonstrated. Proposed enhancements are vague or unfeasible.

Required Knowledge and Skills

- **Understanding of WiFi Protocols:** Knowledge of the 802.11 WiFi standard, including packet structure and transmission mechanisms, and the ability to think creatively about protocol improvements.
- **Wireshark Proficiency:** Familiarity with Wireshark or similar tools for WiFi network data capture and analysis.
- **Basic Python Programming:** Ability to use Python for data analysis and machine learning, including libraries such as Pandas, Scikit-learn, and Matplotlib.

Important Notes

- You are encouraged to use the Ed forum and the dedicated help sessions provided by the tutors.
- Adopt a passive approach to data collection and limit your data collection to data that do not have any user privacy issues associated with it.
- Late penalties will apply at the rate of 5% per day late beyond the due date of submission.

End of Project Specs