Comparison of Content Between "PQClass: Classification of Post-Quantum Encryption Applications in Internet Traffic" to "PQBench: Post-Quantum Encrypted Network Traffic Classification Benchmarking"

Hereby follows a listing of all additional content included in this the manuscript that wasn't included in "PQClass: Classification of Post-Quantum Encryption Applications in Internet Traffic", as detailed in the cover letter:

1. Extended classification experiments:

    A. Key Encapsulation Mechanism classification: We added a new classification task, classifying between the currently deployed Key Encapsulation Mechanism algorithms ML-KEM and Kyber. We conducted two experiments regarding the classification task. The first being differentiate between the two KEMs, and the second is classification of each duo of browser and KEM: Chrome+Kyber, Chrome+ML-KEM, Firefox+Kyber and Firefox+ML-KEM.

    B. POB experiment addition: In addition to the binary PQC, browser, operating system and tuple experiments detailed in the "PQClass" paper, we've also added a new classification task, performing the binary PQC classification task with more packets in each sample. This improved our results for this task.

2. Dataset Enlargement:

    A. New Dataset Introduction – PQC-Algo: For the Key Encapsulation Mechanism classification, we create a new dataset called PQC-Algo. It contains 2400 samples and is divided into four classes detailed in section 1A. The data was collected in 4 different sites over two different countries in various network conditions to ensure variety and real-life implications.

    B. Existing Dataset Extention: We increased the size of the dataset described in the "PQClass" paper, now entitled PQC-POB. Like PQC-Algo, it was collected in 4 different sites over two different countries in various network conditions. The increase in the number of samples has led to improvements in the accuracy of the binary PQC, browser, operating system and tuple experiments.

3. Additional figures and statistics:

    A. Introduction of new measurements: we have added new figures showcasing the relation between the training sample set size to theF1-Score value, to showcase an optimal dataset size needed for successful classification. From that, we took our best-performing model and measured the relation between the training sample set size and classification evaluation metrics, demonstrating how the increase in the training size affects the model.

    B. Extension of existing figures: we have extended the "packets to accuracy" figure from the "PQClass" paper to include 25 and 30 packets per flow, aiming to check the impact such enlargement has on the classification accuracy.