

基於機器學習的加強網路流量分析

Machine Learning-Based Enhanced Network Traffic Analysis

指導教授: 張燕光

專題成員: 簡維頡 黃聖民

開發工具: g++、CMake、gdb、Python

測試環境: Ubuntu 24.04

一、簡介

在傳統的網路流量估測中，Sketch 是一種常見且高效的 hash-based 資料結構，廣泛應用於記憶體受限環境下的即時流量監測。其中 CM Sketch (Count-Min Sketch) 與 CU Sketch (Count-Update Sketch) 為最常使用的兩種變體。然而，由於這類資料結構高度依賴 Hash 函數，容易造成流量高估 (Collision 問題)，進而影響整體估測準確度。

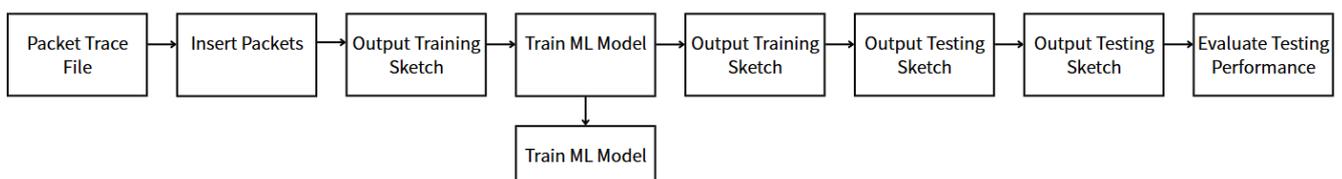
為了解決上述問題，我們參考並實作了《Enhanced Machine Learning Sketches for Network Measurements》中提出的架構，建構出一套結合傳統 Sketch 與機器學習的流量預測系統。本系統以傳統 Sketch 為基礎，透過特徵擷取與監督式學習模型強化預測準確度，實現更加精準的流量預測。

本系統支援兩種預測任務：

Flow Size Estimation：預測單一 flow 在一段時間內的封包數量。

Top-K Flow Detection：判斷在某時間區間內，流量排名前 K 名的 flows。

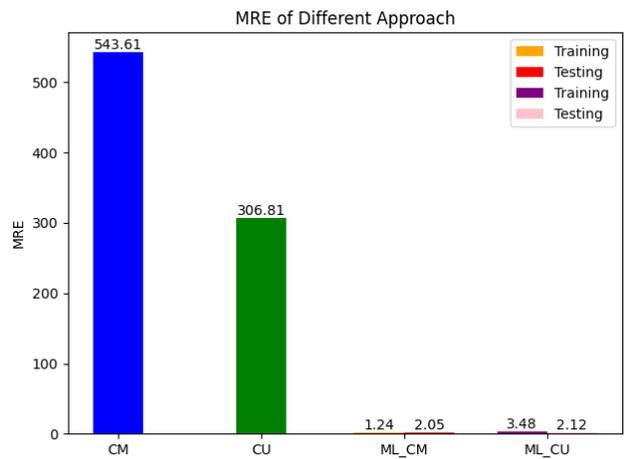
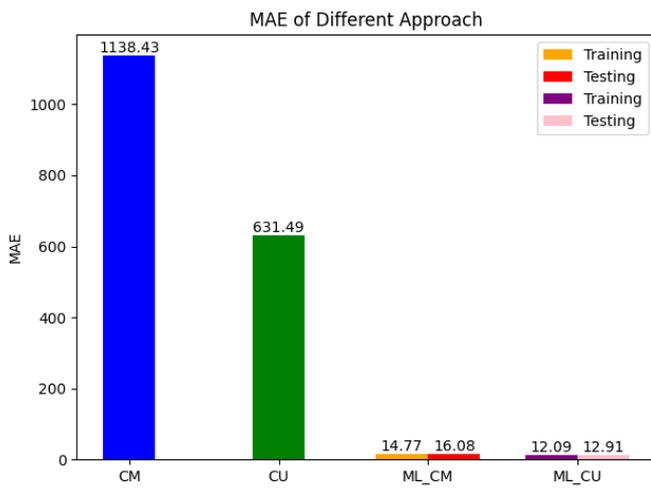
下圖為本系統之整體流程架構圖：



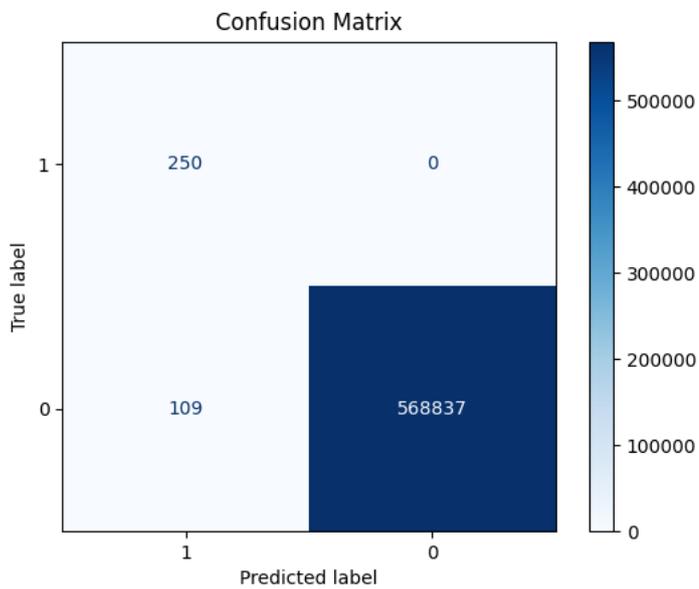
二、測試結果

相較於傳統的 Sketch，我們達到了相較於傳統 Sketch 而言極為優秀的表現，以下為圖表：

Flow Size:



TopK:



Precision@k = 82.4%

Recall = 100%

Accuracy = 99.8%

Precision = 69.6%

F1 Score = 82.1%