中原大學
Chung Yuan Christian University

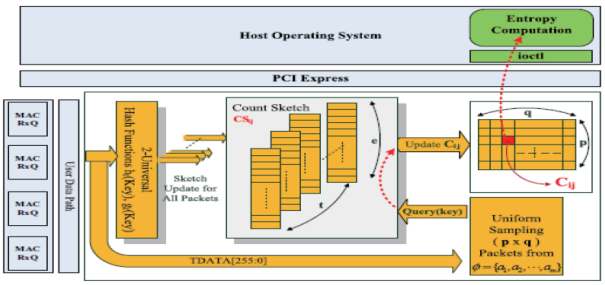
估計熵值之方法與系統

市場簡介

近年來網路發展快速，傳輸速率日益增加，網路的攻擊或是異常行為也越來越普遍，像是蠕蟲(Worm) 、端口掃描(Port Scan)、分散式阻斷服務攻擊(DDoS, Distributed-Denial-of-Service)、位址掃描(Address Scan)等攻擊，都有可能影響到正常的網路環境與使用者。在網路的觀測中，我們可以根據封包標頭資訊進行熵值的統計與分析，用來觀察網路是否有異常的發生。

技術簡介

本發明根據 Ashwin Lall 等人所提出的川流估計熵演算法，使用速寫演算法取代其精確計數的部分，並結合蓄水池取樣演算法，在未知長度的川流資料中以一次性的處理手法，快速的統計封包流的資訊，本發明使用少量的記憶體空間，於 NetFPGA-10G 開發平台，實現基於 Count Sketch 之川流估計熵演算法量測系統。本系統可於 30Gbps 網路流量中，進行高速熵值估算。



在NetFPGA平台上基於熵計算的Count

技術優勢

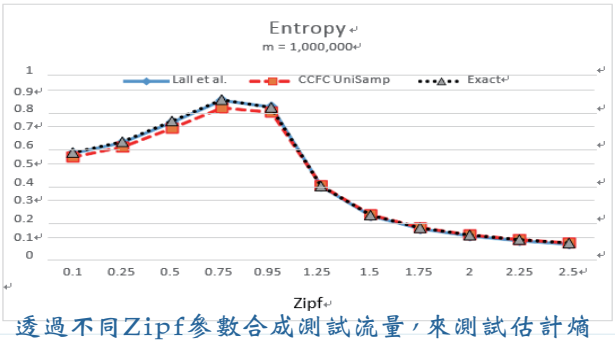
本發明使用相同記憶體空間與較少之 Count Sketch 陣列，比較熵值預估之錯誤率。使用 Count Sketch 演算法，相較 A.Lall 等人所提出演算法，以 4k Byte 為例，我們只需使用近 1/4 的記憶體空間其錯誤率在 5% 範圍內，並可大幅減少系統的運算時間。

	川流估計熵演算法	Count Sketch	Count Sketch(4k)
估計熵空間使用(Kbit)	4kX4 X 200= 3,200	4k X 4 X 32= 512	4k X 4 X= 512
Count Sketch 空間使用(Kbit)	N/A	28k X 3 X 21= 2,688	4k X 3 X 32= 384
總使用空間 (KByte)	400	400	112
錯誤率	0.01%	1.35%	4.23%

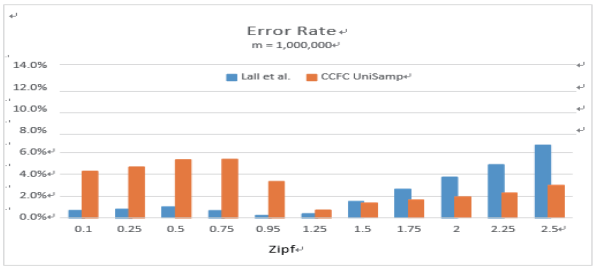
川流估計商演算法空間與錯誤率之比較， z=4k，g=4

項目背景

熵(Entropy)是一種用來量測資料分散與聚集程度的指標，熵的值越高表示資料分佈越分散，而熵的值越低表示資料分布較集中。針對網路封包標頭資訊進行熵值的分析，可以從中得知網路流量分佈的改變，進而找出是否存在異常行為或攻擊的事件。



透過不同Zipf參數合成測試流量，來測試估計熵值。將熵值標準化 後比較三種不同演算法的準確程度。Zipf參數的範圍0.1~2.5，合成數據包含1,000,000封包。



不同Zipf參數的熵值相對誤差，如圖所示，與Lall等人的結果相比，當Zipf參數小於1.25時 Count Sketch方法具有較高的估計誤差，對於大於1.5的Zipf參數，所提出的方法實現較低的誤差。

專利狀態

台灣發明專利→專利號(I541662)

美國發明專利→專利號(US 9,781,427 B2)

合作方式

產學合作

技術轉移

共同合作研發

EOOIA

中原大學產學合作暨專利技轉中心

03-2651831~7

Shun0210@cycu.edu.tw