**Leveraging Traffic Repetitions for High-Speed Deep Packet Inspection**

Anat Bremler-Barr, Shimrit Tzur David, Yotam Harchol, David Hay

IEEE INFOCOM 2015

Summary

1. 此方法是base on AC演算法，主要是收集traffic中出現多次的string，將這些popular string建立成dictionary，透過事先用dictionary的string來search AC，知道search這些string最後會走到AC哪個state，將來在search AC時可以跳過中間幾個state，直接到最後的state，但是不知道要如何定義popular string
2. 因為dictionary的string是用hash table來儲存的，所以如果一個gram是8 bytes，一開始就要讀8 bytes(令 bi ~ bi+7)做hash，看hash table中是否有這個字串

Case1: dictionary miss

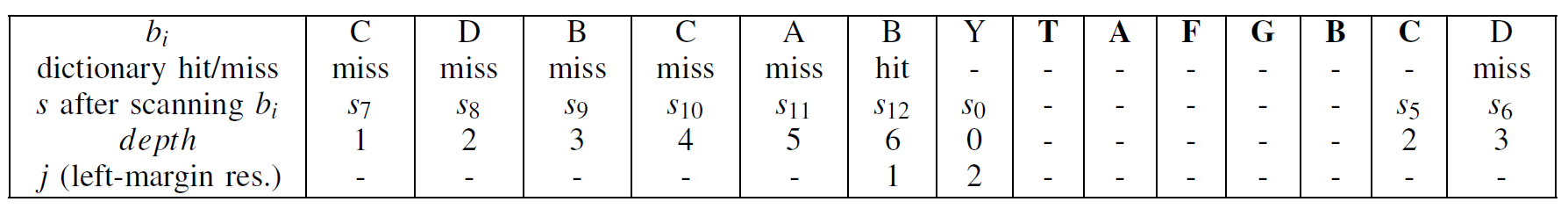
就用bi做search，紀錄走到哪個state和depth，接著讀下一個8 bytes

Case2: dictionary hit

j是這個character在gram中的位置

如果depth >= j 代表這個gram可能會match長度更長的pattern，所以必須要讀下一個character

如果depth < j 代表search這個gram的過程中失敗了，所以回到state 0或者經由failure link走到深度更淺的state，這時就可以跳過這個gram，直接走到最後的state，接著讀下一個8 bytes



1. Paper有計算查詢Bloom filter、AC和dictionary的時間，p是可以跳過這個gram的機率，pmin是p要比pmin的值還大，這個方法才會比原本的AC還快

這方法因為要收集traffic來建立dictionary，可是paper並沒有將建立dictionary的時間考慮進去

dictionary每隔一段時間就要update，paper也沒有考慮進去

