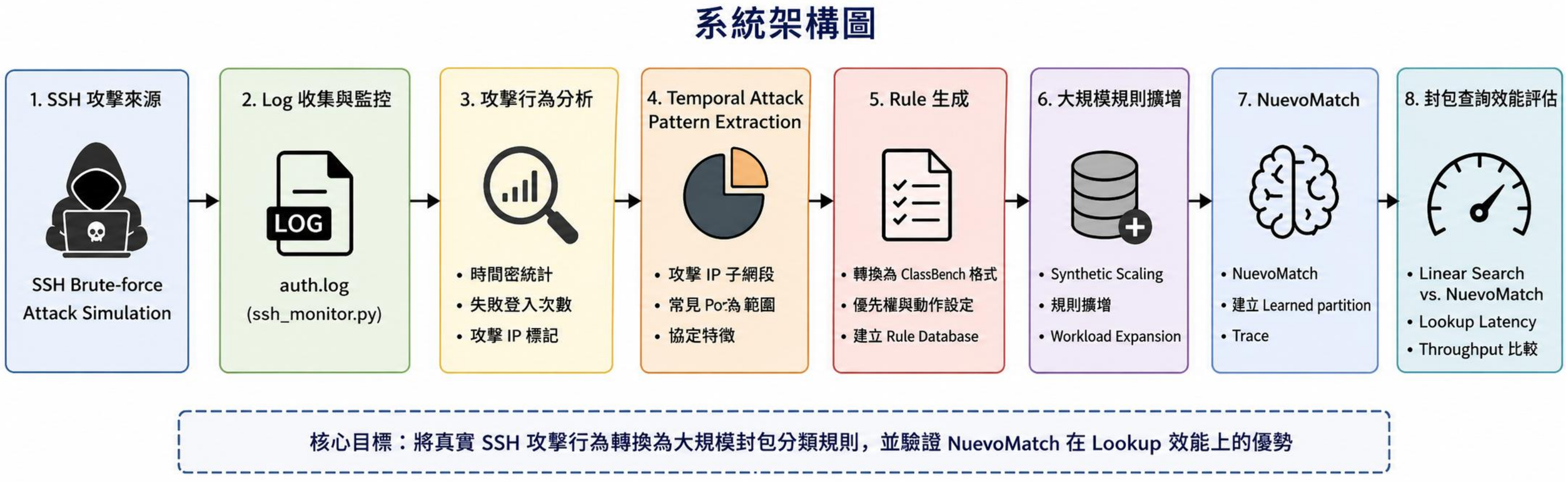
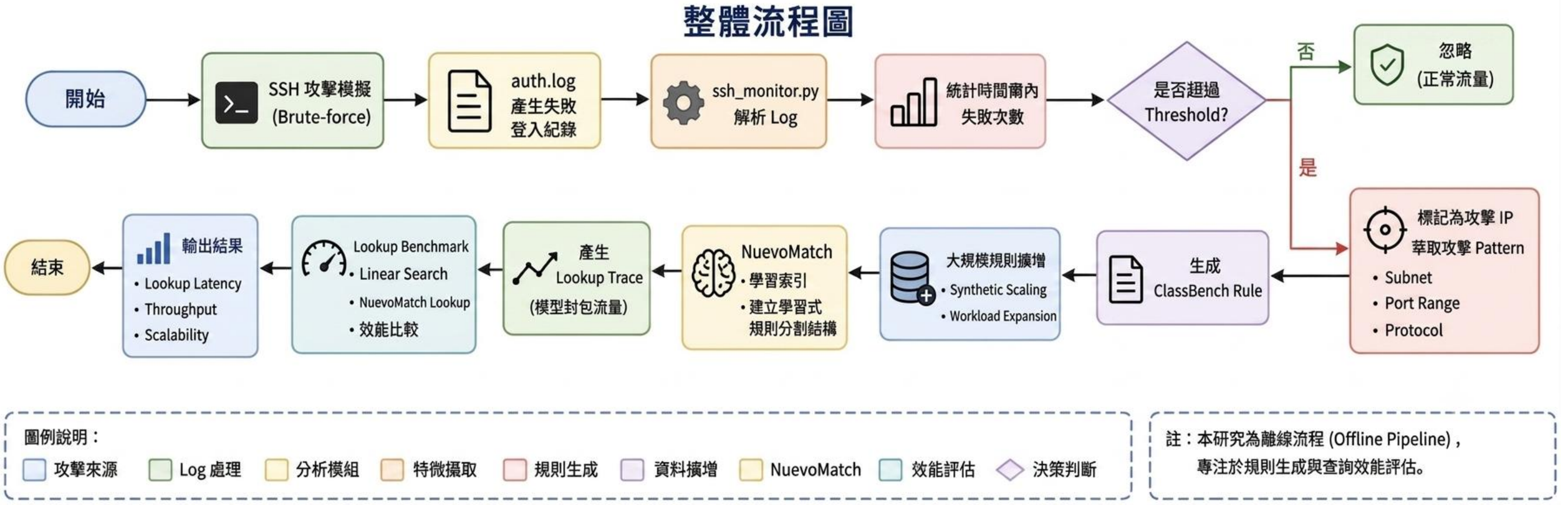
基於 RQ-RMI 學習型索引之自適應封包分類系統實作

現在的封包分類相關研究有非常多種，大部分都是使用靜態 ClassBench 來進行資料測試，但是規則的內容和真正的網路攻擊行為始終缺乏一套完整的流程。因為真正的網路攻擊中很難找出封包查詢的特徵。所以我們嘗試把模擬的SSH暴力攻擊記錄都記錄至日誌中，最後再把他變成封包學習分類系統可以使用的規則，再進一步的檢驗這樣的流程再大量規則查詢時他的效能表現。

我選擇了NuevoMatch來當我們實作的主架構，他是一種混合式封包分類架構，其中包含了 learned partition（iSets）、基於回歸的學習索引模型(類似RMI架構)以及傳統 Cutsplit和EffiCuts等分類器做回退機制(fallback)，以提升大規模規則查詢效能。並搭配 trace generator 產生封包流量進行 lookup benchmark。並把之前所生成出來的 ClassBench 規則集拿來建立學習式的規則分割結構，透過學習模型來預估可能命中的規則分布區間，進而降低傳統線性搜尋的成本，因為傳統線性搜尋需要針對規則做逐條比對，其成本可相當可怕。





學習式封包分類和傳統線性搜尋方法比較結果中，我們發現規則集從50K成長到500K的時候，雖然NuevoMatch的封包查找時間也會緩慢上升的趨勢，但仍然維持在微秒等級，反觀傳統線性搜尋方法的封包查找時間是呈現明顯的線性成長。在Benchmark實作結果中可以發現到NuevoMatch在500K的規則集下平均查詢延遲時間為3.336 μs/packet，吞吐量大約是299,730.24 packets/sec。而線性搜尋方法在500K規則集的時候其查找時間已經高達數百秒了，因此我們得出的結論是傳統線性搜尋方法在大型規則集下將會面臨擴展性的這個瓶頸。與此同時我們也從結果中發現學習式封包分類透過學習索引和錯誤邊界搜尋(Error-bound Local Search)這方法可有效的降低大規則集下的封包查找時間成本，並且在擴展性上也有較好的表現。

另外我們也發現到在大型規則集擴展的過程中可能會出現萬用字元重疊(Wildcard overlap)、優先級遮蔽(Priority shadowing)、規則永遠不會被命中(Unreachable rule)等現象。雖然我們在規則集的擴展是以模擬SSH攻擊的規則和合成式規則擴增方法(Synthetic Rule Expansion)結合，但此結果也反應出在真實的大型封包分類流程中像是大型資料交換中心一樣會發生的現象。在實做中我們除了驗證學習式封包分類在封包查找的效能外，也模擬了在大型ACL環境中可能出現的規則重疊與遮蔽現象。