

# 基於Hadoop平台之APT攻擊大數據分析研究

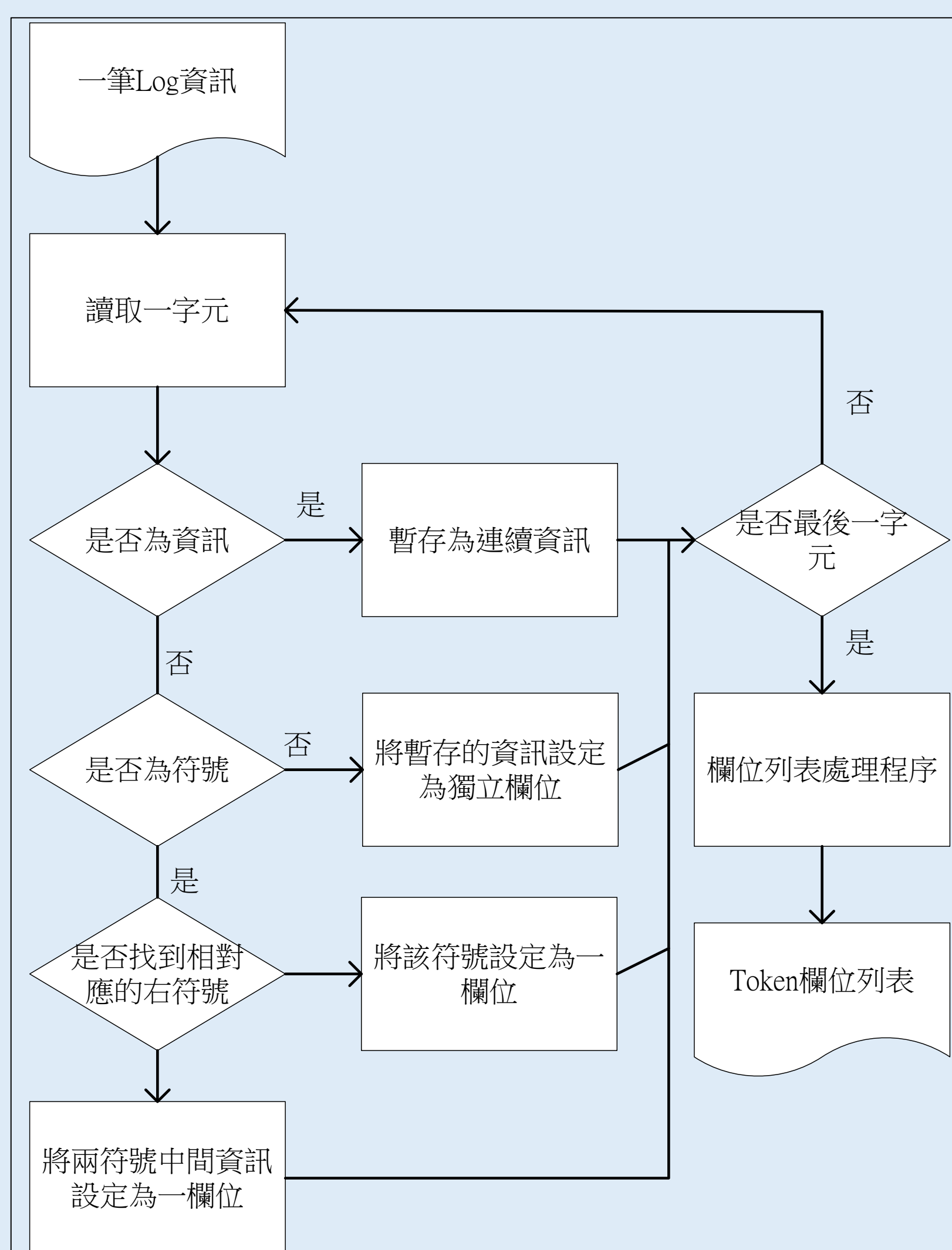
計畫主持人:賴谷鑫

執行單位:臺灣警察專科學校

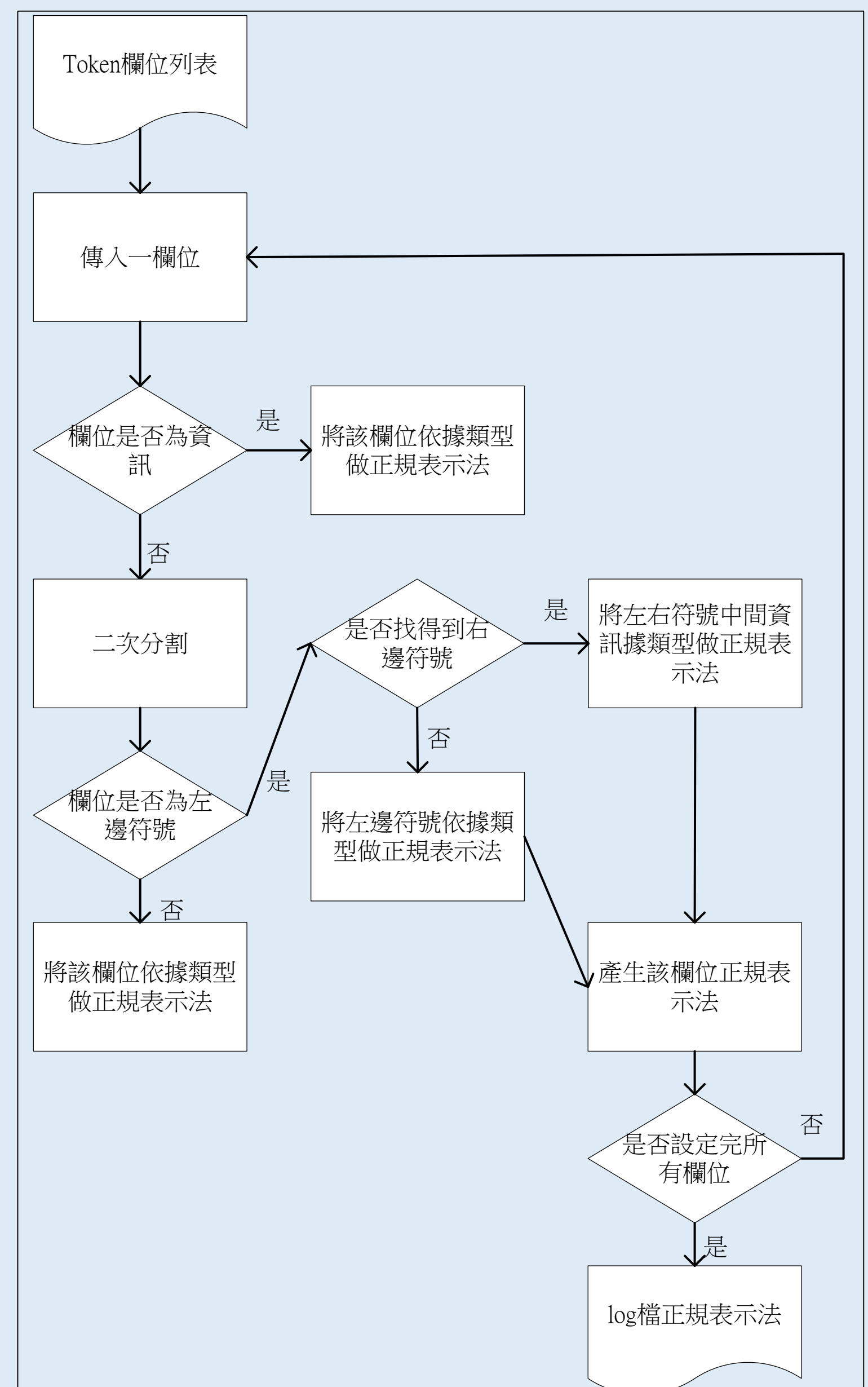
## 簡介

SIEM藉由整合以及分析組織內之資訊安全設備日誌以偵測APT攻擊，然而SOC在處理資訊安全的紀錄檔時面臨第一個挑戰即為如何將異質資料格式進行正規化，進而使後端的分析平台可以根據這些資訊分析、告警以及產出報表。有鑑於此，本研究開發一套智慧型記錄檔剖析系統（Intelligent Parsing System）透過該系統可以針對不同日誌快速產生高品質之正規表示式；另本研究更以Hadoop以及Spark為基礎架設雲端分析平台，以整合、關聯與分析這些經過正規化之資安設備日誌。

## Token 拆解流程圖



## 正規表示式產生流程圖



## 系統網站

The screenshot shows the 'Log Transfer' web interface. It includes a text input field for a log entry, buttons for 'ArcSight Transfer', 'Logstash Transfer', 'Fluentd Transfer', and 'Upload File'. Below this is 'Step 2: 設定Token資訊', which displays a table with 4 tokens. The table has columns for Token Name, Token Type, Token Data, and Function. The tokens are: '192.168.3.111', '08/Mar/2015:03:45:10 +0800', 'POST /api/ HTTP/1.1', and 'User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; InfoPath.1; .NET CLR 2.0.50727)'. Below the table is 'Step 3: 自訂正規表示法(非必要)', which shows a 'Log Regex' field with a generated regex and a 'Generate Regex' button. At the bottom, there are buttons for 'Set Variables', 'Set Submessage', and 'Generate'.

Token Name	Token Type	Token Data	Function
	String	192.168.3.111	↑ × ↺
	String	08/Mar/2015:03:45:10 +0800	↑ × ↺
	String	POST /api/ HTTP/1.1	↑ × ↺
	String	User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows	↑ × ↺