大型校園開放式服務系統
異常帳號偵測機制介紹

國立臺灣大學計資中心網路組 邵喻美

July 2023

Outline

- 帳號、服務之盜用情況及影響
- 異常登入偵測機制
- 異常登入偵測實作-ELK
- 實際操作案例

服務盜用情況分析

臺大VPN服務

- 在中國翻牆
- 掩蓋真實來源以登入臺大服務
- 大量下載圖書館電子期刊

➢佔用使用授權&浪費網路資源

➢大量下載電子期刊

➢當作跳板入侵/攻擊其他目的端

➢危及帳號安全 → 曾有盜用者冒名回信

計中帳號、信箱

- 盜用信箱寄發垃圾/釣魚郵件
- 測試登入收集帳號密碼
- 長期潛伏收集情資

➢導致臺大mail server被列入黑名單，寄信被拒絕

➢可能影響被盜帳號收發信件

- ➢修改寄件者名稱
- ➢設定「收件匣規則」→所有收到信件一律刪除

Email threats

- Incoming threats
  - 惡意郵件：spam/virus/phishing
  - 流向：從校外寄進來、校內信箱被盜於系統內寄信
  - ➢ 影響：導致使用者電腦中毒、資料被竊、帳號被盜等損害
  - ✓ 對策：
    - ✓ anti-SPAM/anti-Virus + AMP(Advanced Malware Protection)
    - ✓ Ironport + Exchange server default/customed Filter
- Outgoing – threats from within
  - 從被盜帳號向外部發出垃圾/惡意郵件
  - ➢ 影響：危害本校郵件伺服器名聲，導致對外寄信被拒
  - ✓ 對策：”**及早**”偵測異常帳號

帳號與服務

類型＆狀態 \ 服務		數量	VPN	Mail2.0	Mail1.0
教職員	在職＆退休	約15000個	○	○	╳
公務計畫帳號	在職	約780個	╳	○	╳
學生	在學＆休學	約2萬個	○	⊙	⊙
	畢業	超過20萬個	╳	⊙	⊙
校友			╳	╳	○

服務特性

- **VPN**
  - 每個帳號同時只可一次登入，每日不限登入次數
- **Mail 2.0**（不限登入地點）
  - 支援協定：SMTP、POP3、IMAP4、Exchange、OWA、ActiveSync
- Mail 1.0（不限登入地點）
  - 支援協定：SMTP、POP3

➢ **上述服務皆不限來源**

異常登入偵測機制

偵測異常帳號

- **從結果判斷**
  - 信箱被盜用大量發送垃圾信件
    - 設定信件量門檻值，定期統計&偵測大量寄信者
- **從跡象判斷**
  - 異常登入
    - 定期統計短期間內從某數量以上的不同國家IP登入者
    - 從數個不同國家IP位址登入計中服務（同一服務或不同服務，如VPN、Mail）
    - 使用過去曾出現的可疑 mail agent/VPN hostname
    - 與**過去**使用型態不同的登入方式 → Trend
      - 從未/很少使用該服務、從很少見的國家登入

異常登入情境

- **同一帳號一天內從不同國家登入使用不同服務**
  - 彙整**VPN登入記錄**及**各郵件協定log**，比較同一帳號登入相同及不同服務的記錄
  - 可偵測不一致的登入行為，例如
    - 從臺灣收發email，但從中國登入VPN
    - 透過OWA（Web/手機）從臺灣收發email，但SMTP/IMAP4從其他國家登入（未必寄信）
  - 例外：
    - 在國外使用VPN軟體，會出現從多個國家連線的現象（可從設備名稱判斷為同一設備）
    - **同時從臺灣及國外收發email**：透過其他mail service收信，如gmail、microsoft等
      - 從設備名稱判斷，包含Outlook、行動裝置（smart phone & pad）➔ 盜用者很少用行動裝置登入
    - **無VPN登入紀錄卻以VPN IP登入信箱**：以在學生帳號登入VPN後再登入已畢業漲號信箱、以職員帳號登入VPN後再登入公務漲號信箱
- **帳號從單一「高風險」國家登入**
  - 將各郵件協定登入來源國家依帳號數量Desc排序
  - 目標：從罕見/少用國家找可疑登入
  - 例外
    - 使用者真的出國 → 從使用者是否有使用該郵件協定的習慣，前後幾天的記錄，以及寄信是否正常等跡象判斷
    - 難以判斷從臺灣IP及其他「正常」國家的登入記錄，或者夾雜在正常記錄裡
    - Anonymous VPN越來越普及，可從使用者使用設備及習慣判斷

異常登入偵測方式

- **定期統計各帳號在一段期間內登入SMTP/IMAP4協定的IP/as.org數量**
  - 統計SMTP/IMAP4 log各帳號登入IP數量，排除臺大、google、*.tw、*.com等domain（**Script自動偵測及封鎖）（透過 ELK 呈現）**
  - 可偵測到寄信量未超過門檻值或甚至未寄信的異常登入行為
- **同一帳號一天內從不同國家登入使用不同服務**
  - 彙整各郵件協定及VPN log，比較同一帳號登入相同及不同服務的記錄**（利用 ELK 呈現&搜尋）**
  - 可偵測不一致的登入行為，例如
    - 透過OWA（Web/手機）從臺灣收發email，但SMTP從其他國家登入（未必寄信）
    - 從臺灣收發email，但從中國登入VPN
- **帳號從單一「高風險」國家登入**
  - 將登入來源國家依帳號數量Desc排序**（利用 ELK 呈現&搜尋）**
  - 可偵測盜用帳號者利用各種協定測試登入

偵測目標－出現新的服務型態

- **偵測帳號是否出現過去幾個月不曾使用過的服務型態**
  - 目的：利用盜用者測試/登入時的協定可能與使用者平時使用型態不同作為判斷條件（**程式定期讀取ES資料比對**）
  - 例外：新使用者會誤觸條件(尤其每年學期初)
  - 做法：
    - 根據已匯入ES的帳號登入紀錄進行統計
    - 每天將帳號使用的服務型態，與從ES中撈出該帳號過去的使用型態比較
    - 每日產出報告寄給管理者確認

異常登入偵測實作

運用ELK彙整、呈現、比較資料

處理流程 – 1. 處理raw log

目標：擷取出對於判斷帳號登入有意義的資訊

- **VPN log**：撰寫程式分析access log，擷取出每個帳號連線資訊（每天約15萬行log）

服務	日期	時間	帳號	VPN設備:分配IP	來源IP	hostname:user agent

- **POP3/SMTP/IMAP4 log**：撰寫powershell程式，利用微軟logparser程式解析帳號登入資訊（每種協定一天約61個檔案x4台伺服器）
  - 此三種簡單的郵件協定沒有user端如user agent等相關資訊

服務	日期	時間	帳號	ServerIP	來源IP

- **Exchange W3SVC1 log**：撰寫powershell程式parse logfile內容，根據多項欄位參數內容判斷屬於OWA/Exchange/ActiveSync協定，並擷取登入資訊（每天約7~9Gx4台伺服器）
  - 每個 W3SVC1 檔案從GMT 00:00 開始/結束，所以每天上午8點後開始處理前一天紀錄

服務	日期	時間	帳號	Server IP	來源IP	User agent

- **MessageTracking Log**：撰寫powershell程式，利用微軟logparser程式解析帳號寄信紀錄（每天約24Mx12台伺服器）
  - 統計各信箱透過mail server寄信的紀錄(不含外部寄信進來的紀錄)

日期	時間	Client IP	Server IP	sender	Message subject	recipients

處理流程－2. 彙整資料

目標：整合各項服務紀錄、formalize data

- 將所有初步擷取出來的紀錄檔彙整到同一台server以便處理
- 撰寫shell及perl程式
  - 處理windows系統檔案格式，並將各紀錄檔整合成一個檔案
  - 排除不需要及重複的資料，例如【帳號+服務+來源IP+server IP+user agent】完全一樣的資料
  - ~~將來源IP轉換成geo_location~~ ➔ 已改為匯入ELK之後直接使用geo_ip
  - 處理寄信紀錄檔，分別計算各帳號從校內IP及從校外IP寄信數量

處理流程 – 3. 匯入ELK

目標：

1. 利用 ELK Dashboard 檢視/比對/查詢 所有服務登入紀錄
2. 直接讀取存放於ElasticSearch的紀錄進行使用型態比對，產出每日新登入服務報告

- 每天定時將處理過的log資料匯入ELK
- 利用shell/python程式每天定時擷取網路組各帳號前一天所有服務登入紀錄寄到個人信箱
- 利用python撰寫程式：從ES讀取前一天各帳號登入的服務資訊，再與該帳號在ES中所有登入服務比較
  - 出現新服務便記錄下來，寄給管理者
  - 出現新國家 ➔ 排除大宗ISP後資料量仍過多，待優化

實際操作案例

常見異常登入帳號偵測型態

- 日常作業 → 每日登入報告&ELK介面瀏覽
- 出現新使用型態 → 每日登入報告
- 交叉比對搜尋相同類型異常登入 → ELK介面

偵測異常登入機制的優勢

- 在盜用者測試或潛伏登入時就掌握狀況，提早提醒使用者修改密碼
- 從類似型態一舉揪出一批異常帳號
- 多次阻攔下帳號被盜用後大量對外寄信的情況
- 建立可疑型態資料庫，例如as_org、agent等，平日檢視ELK時可依據先前可疑條件搜尋是否出現異常型態

經驗檢討

- 瞭解你要的資料
  - 分析各類型日誌檔內容意義
  - Session型態日誌檔的處理
  - 版本更新後log格式更動
- 資料前處理
  - 格式formalization
  - 撰寫程式處理 vs. 全部匯入Logstash
- Get what you need
  - 由瞭解需求的人設計視覺化界面
- 準確性 vs. 可用性
  - 「寧可錯放」還是「寧可錯殺」？

建議與討論