Regular Paper

Communications of the CCISA
Vol. 31 No. 3 Aug. 2025

AI 方法建構 APT 駭客家族攻擊情資特徵拓樸及風險評估架構-以 AD 伺服器攻擊為案例研究

王仁甫[1*]、陳品翰[2]、陳思宥[3]、呂書晴[4]、徐睿廷[5]、陳品嘉[6]、魏得恩[7]
[1,2,3,4,5,6,7]元智大學資訊管理學系
[1]Fisher2023@saturn.yzu.edu.tw、[2]s1146212@mail.yzu.edu.tw、
[3]s1101651@mail.yzu.edu.tw、[4]s1111615@mail.yzu.edu.tw、[5]s1121730@mail.yzu.edu.tw、
[6]s1121738@mail.yzu.edu.tw、[7]wilbur.wei@saturn.yzu.edu.tw

摘要

現行進階持續性威脅 (Advanced Persistent Threat, APT) 多以非結構化文本呈現，難以與漏洞資訊有效對應，使的風險評估缺乏整合攻擊特徵、攻擊者族群與實際漏洞利用行為的科學模型。本研究提出結合拓樸理論與人工智慧 (Artificial Intelligence, AI) 的 APT 攻擊情資分析框架，以 Active Directory 攻擊為例，蒐集五年台灣資安防護體系關注的資安情資，建構拓樸圖並應用 PageRank、TrustRank 與 RST 量化風險；同時以 LoRA 微調 LLaMA 3.1 預測 CVSS 指標。結果顯示，此方法能有效辨識 APT 家族行為特徵並建構動態風險評分，提升修補決策精確度。

關鍵字：資安情資、AD 攻擊、拓樸理論、APT 家族

* 通訊作者 (Corresponding author.)

AI-Based Construction of APT Hacker Family Attack Intelligence Topology and Risk Assessment Framework: A Case Study of Active Directory Server Attacks

Jen-Fu Wang[1*], Pin-Han Chen[2], Ssu-Yu Chen[3], Shu-Ching Lu[4], Jui-Ting Hsu[5], Pin-Chia Chen[6], Te-En We[7]
[1,2,3,4,5,6,7]Department of Information Management, Yuan Ze University
[1]Fisher2023@saturn.yzu.edu.tw、[2]s1146212@mail.yzu.edu.tw、[3]s1101651@mail.yzu.edu.tw、[4]s1111615@mail.yzu.edu.tw、[5]s1121730@mail.yzu.edu.tw、[6]s1121738@mail.yzu.edu.tw、[7]wilbur.wei@saturn.yzu.edu.tw

Abstract:

Current Advanced Persistent Threats (APTs) are predominantly presented in unstructured text formats, making it difficult to effectively correlate with vulnerability information. This results in risk assessments lacking scientific models that integrate attack characteristics, attacker groups, and actual exploitation behaviors. This study proposes an APT attack intelligence analysis framework that combines topology theory and Artificial Intelligence (AI). Using Active Directory (AD) attacks as a case study, we collected five years of cyber threat intelligence relevant to Taiwan's cybersecurity defense ecosystem. We constructed topological graphs and applied PageRank, TrustRank, and RST to quantitatively assess risk. Additionally, LoRA fine-tuning of LLaMA 3.1 was used to predict CVSS metrics. The results demonstrate that this approach effectively identifies behavioral characteristics of APT families and constructs dynamic risk scores, thereby improving the accuracy of remediation decision-making.

Keywords: Cyber Threat Intelligence, Active Directory Attacks, Topology Theory, APT Families

壹、前言

資安維運面臨大量情資解析與採取適當防護策略的挑戰；以公開的情資平台 ioc.one 為例，單日釋出逾 600 筆以上情資，卻少有研究台灣資安防護體系關注的攻擊場景。雖然已有研究導入大型語言模型協助情資整理與攻擊手法建模，仍面臨情資量龐大影響模型效能、摘要與特徵提取困難、特殊標籤識別率低等問題，導致情資難以與漏洞資訊結合，降低情資可用性。

因此，本文根據及彙整 Wei et al. (2025) 及王仁甫 (2025) 之論文為基礎，試圖使用 AI 方法論，分析資安漏洞情資風險，更有效的 ZTA 風險評分機制。

貳、文獻探討

2.1 資安漏洞情資風險

自從 2005 年美國資安事件應變組織 FIRST 發布通用漏洞評分系統 (Common Vulnerability Scoring System, CVSS[1]) 後，該系統已成為全球政府與產業評估漏洞風險的標準。然而，面對快速成長的漏洞數量，使得企業多依賴資安專家主觀判斷與 CVSS 分數嚴重程度來決定優先修補順序 (Jacobs et al., 2021)，導致組織難以全面修補所有漏洞，進而讓駭客有入侵的機會。

依照 CVSS 指標規範，漏洞概念由八個子概念組成[2]，但仍難以衡量資安風險，因此 Syed (2020) 進一步提出四個補充子概念: 漏洞嚴重性(基本評分為 0 到 10，以及相關的定性嚴重程度分級)、可利用性(漏洞利用的次數)、類型(CVE 8 提供的漏洞類型)，以及揭露情形(如是否延遲揭露與揭露日期)，以提高漏洞風險評估的完整性。

然而，CVSS 主要衡量漏洞「嚴重性」，而非實際「風險」。因此，FIRST 於 2015 年提出漏洞預測評分系統 (The Exploit Prediction Scoring System, EPSS)，針對每個 CVE 漏洞預測其 30 日內被利用的機率，並每日更新。EPSS 考量計算成本與覆蓋率，協助企業聚焦修補具實際利用風險的漏洞，提升資源分配效益。

此外，Jacobs et al. (2021) 發展一套漏洞風險評估框架與 EPSS 模型，透過主成分分析 (Principal components analysis, PCA) 選出 16 個變數，建構供應商 (Vender)、可用性 (Exploit) 與漏洞特徵三大指標類別 (Tag)。研究利用邏輯迴歸預測 12 個月內漏洞被利用的可能性，結果發現某些供應商 (如 Google) 雖通報漏洞數量多，但實際被利用比例較低；而 CVE 中引用文獻數與漏洞利用機率呈正相關。

---

[1] 1999 年 CVE 計畫啟動後，由 MITRE 公司執行，經費來源為美國國土安全部 (DHS) 和網路安全和基礎設施安全局 (CISA)。詳見:NIST 說明: https://nvd.nist.gov/general/cve-process；及 FIRST EPSS 模型: https://www.first.org/cvss/v1/intro(最後查詢日:2025 年 4 月 2 日)。

[2] 攻擊媒介（漏洞可能被利用的途徑）、攻擊複雜度（漏洞是否容易被利用）、所需權限（攻擊者在利用漏洞前需要具備什麼樣的系統權限）、使用者互動（是否需要受害者執行某些操作來觸發漏洞）、機密性影響(漏洞是否會導致敏感資料外洩)、完整性影響(資料是否會被未授權修改)、可用性影響(系統是否可能無法正常使用)

2.2 APT 攻擊情資特徵

於資安實務面上，當工程師接收到關於進階持續性威脅組織(以下簡稱 APT 駭客族群)的攻擊情資時，需依賴專業判斷分析 IP、網域與漏洞資訊，評估其與內部系統的關聯性，進而確認資安風險。然而由於情資碎片化，分析過程需在短時間內整合大量資訊，易導致判斷失誤；此外，專家主觀分析可能產生偏見或代理人問題，難以因應情資與事件的快速增長。因此，企業應結合黑名單情資和受害者相似度量模型，以提升風險評估的準確性 (Rezapour et al., 2020)。

在此背景下，APT 攻擊情資特徵對於企業之重要性也隨之提升。而現有文獻對於 APT 家族之攻擊主要的研究大多集中於對其性質 (Ussath et al., 2016)、生命週期 (Vukalovic & Delija, 2015) 及網路行為 (Lamprakis et al., 2017) 的特徵分析。Liras et al. (2021) 則從惡意軟體中提取動態及靜態特徵，以進行 APT 攻擊家族識別。Li et al. (2024) 進一步提出整合 APT 攻擊對象與行為特徵的方法，建構 APT 組織間的網路拓撲結構，並以節點相似度模型刻畫駭客行為。該研究以 APT 29 和 CosmicDuke 做為案例進行驗證，從拓譜圖發現 2016 年到 2018 年期間，APT 駭客族群之間的關係變得更加混亂；然而在 2019 年至 2022 年間，大多數 APT 組織都形成了自己獨特的攻擊模式，關係趨於穩定。

然而，目前文獻多聚焦於 APT 駭客族群的惡意程式、攻擊指標 (Indicators of Compromise, IOC) 與攻擊策略 (Tactics, Techniques and Procedures, TTPs) 之比較，對於 APT 家族所利用的漏洞、IOC 與實際事件之間的關聯性卻少有深入探討，凸顯了情資分析中資訊混雜的問題，影響情資的可用性。

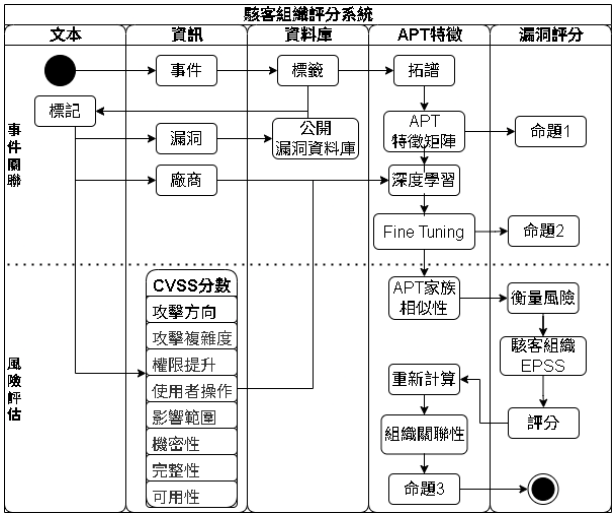
因此，本研究將嘗試使用拓譜理論及 AI 的相關方法，萃取 APT 駭客族群關鍵情資，進而建立漏洞被利用性的風險評分機制以提升企業面對 APT 威脅的風險管理效能。

參、方法

本研究為研究台灣資安防護體系關注的資安情資，故建構爬蟲系統，蒐集 5 年多的中文情資文本[3]，針對台灣論壇關注攻擊事件進行標記[4]，包含漏洞利用、駭客家族、惡意程式、攻擊特徵、受害組織類型，接著使用拓樸理論 (Topology theory) 建立關聯向量，再整合 FIRST 等單位所發布 CVSS 及 ESS 等漏洞評分資料，進一步建置駭客組織評分系統資料庫(詳見下圖一)。

[3] 收集 5 年多(2018 年至 2024 年 3 月)的 IThome 中與資訊安全中文情資相關的文章，經清理資料(如有廣告符號的標記)

[4] 中研院字庫進行斷詞、斷句後，再對應文獻中的特徵進行整理。



圖一：研究架構

再者，本研究先使用 Apriori 關聯演算法建構關聯拓譜，並參考相關文獻採用 PageRank 與 TrustRank 等圖論方法，分析 APT 攻擊情資特徵，進而萃取不同駭客家族的特徵矩陣，以確認研究**命題 1： 高風險 (CVSS 評分高)的漏洞，於資安事件群中的拓譜中，所具有的關聯性。**CVSS 雖能描述漏洞基本屬性，但需結合資料驅動的威脅評估 (如 EPSS) 使用，以便提供組織有效判斷修補優先順序。

第二階段，本研究使用 Llama 3.1 模型進行訓練，分析資料庫中漏洞的利用率等指標，並導入約略集合理論 (Rough Set Theory, RST)，區分 APT 駭客族群攻擊情資的關聯特徵與拓撲結構。最後，以 Active Directory 為案例，驗證**命題 2：ATT&CK 的攻擊程序能否有效分辨 APT 駭客族群。**

最後，本研究將上述 APT 駭客族群相似性特徵結合 CVSS、EPSS 等評分機制，研究**命題 3：AI 分析能否建立更有效的風險評分機制。**

主要原因在於，組織通常會優先選擇高風險 (高分) 的漏洞情資進行修補，然而，由於系統老舊或其他因素，修補工作往往難以實現。因此，組織可能會調整 CVSS 中的環境風險分數，而非進行修補，但此舉缺乏完整的科學依據。若能利用 AI 模型提供漏洞修補與資安事件情資的建議，或許能提供更客觀且安全的決策依據。

3.1 拓樸理論 (Topology theory)

本研究參考 Jackson (2008) 所著的<Social and Economic Networks>書中[5]以及相關文獻分析研究 APT 攻擊手法，例如 Amit et al. (2018) 使用蜜罐資料研究會談層的暴力攻擊 (brute force attacks, BFAs)[6]，並應用拓譜理論分析攻擊行為，例如使用中心性、連

[5] Jackson, M. O. (2008). Social and economic networks / Matthew O. Jackson. Princeton University Press.

[6] 當駭客於第一次會談就能夠正確的登入帳號及密碼，代表已經由第三方完成暴力攻擊，破解該入帳號及密碼。

結性等指標了解攻擊來源國。

因此，本研究運用拓樸理論探討駭客家族的(社會)網路結構，將 APT 駭客家族視為主要節點，攻擊手法(如利用漏洞)則為連接節點，並透過鄰接矩陣 (Adjacency Matrix) 描述網路中的連接情況。其中，鄰接矩陣 $A=[a_{ij}]$ 中，若 $a_{ij} = 1$ 則表示節點 i 和 j 之間存在邊；反之則無，此矩陣是分析駭客家族的網路結構特性的基礎工具。

在此基礎上，本研究導入尤拉路徑 (Eulerian Path) 與哈密頓路徑 (Hamiltonian Path) 的概念，以建立 APT 家族的攻擊特徵與路徑。其中，尤拉路徑代表圖中每條邊皆經過一次且僅一次，根據尤拉提出的判定定理，若圖中最多僅有兩個奇數度數節點，則該圖存在尤拉路徑。具體而言，對於具有 V 個節點和 E 條邊的連通圖，若所有節點的度數皆為偶數，則該圖存在尤拉迴路。而哈密頓路徑則是指圖中每個節點僅經過一次的路徑[7]。

本研究利用上述的拓樸理論，依駭客家族的攻擊手法(如使用漏洞)，建立及優化駭客家族間的關聯性，進一步探討網路中節點的連結特性與重要性，分析指標包括：度中心性 (Degree Centrality)、接近中心性 (Closeness Centrality) 與中介中心性 (Betweenness Centrality)。

中心性公式為 $Cc(i) = \sum_j a_{ij}$，表示節點*i*與其他節點的直接連接數量，也就是該節點的度數用以反映該駭客家族與其他攻擊手法或目標間的直接關聯程度。而接近中心性 $Cc(i) = 1 / \sum_j d(i, j)$ 測量節點 *i*到其他節點的平均最短路徑距離，數值越高，表示該家族常用的攻擊手法與腳本[8]在整體攻擊關係中更具代表性與效率。

最後，我們會探討網絡的全局特徵，如直徑 (Diameter)、平均路徑長度 (Average Path Length)、和聚類係數 (Clustering Coefficient)，以衡量 APT 駭客家族間的關聯程度。

其中，聚類係數的公式為: $C = (3 \times \text{number of triangles}) / (\text{number of connected triples})$，用於量化網絡中節點間形成三角形連接的程度。

3.2 Apriori 關聯演算法

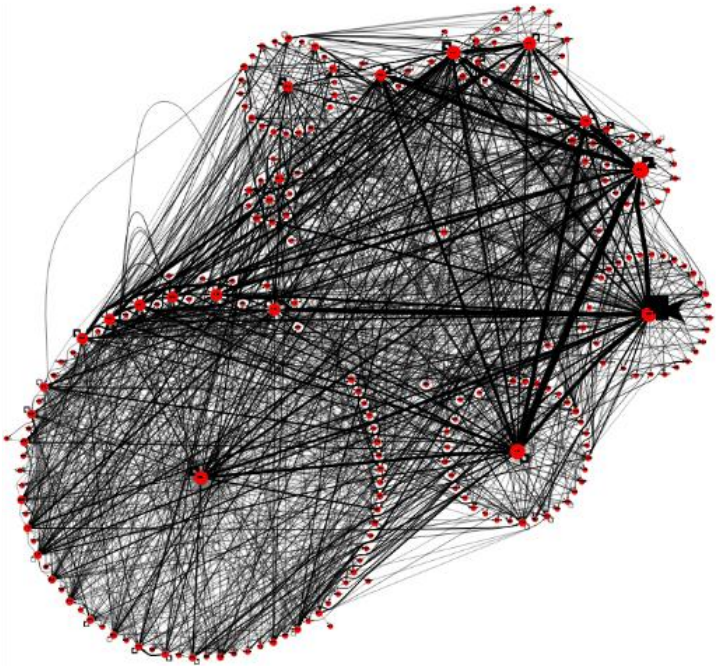
本研究使用 5 年多的中文情資文本[9]，研究台灣論壇關注攻擊事件特徵、拓譜關聯及漏洞利用，再使用 AI 介接分析漏洞利用率等指標分析後，使用中研院字庫進行斷詞、斷句，進一步繪製初版 Key Graph (如圖二所示) 。

由圖二中可了解，強連結的圓圈中心點代表主要的屬性為水準核心價值，透過這些核心價值與其他名詞之間的關聯；而圖中的連結是使用 Apriori 關聯演算法，將字詞之間的關係以 key-value 的組合建立關聯。

[7] 雖然哈密頓路徑的存在條件沒有像尤拉路徑那樣簡單的判定方法，但它能解旅行推銷員問題 (Traveling Salesman Problem, 簡稱 TSP)，以最佳化駭客家族的網路結構。

[8] 此外，中介中心性則反映某節點在最短路徑中出現的頻率，是衡量該節點在網絡中橋樑作用的指標。

[9] 收集 5 年多(2018 年至 2024 年 3 月)的 IThome 中與資訊安全中文情資相關的文章，經清理資料(如有廣告符號的標記)



圖二：資訊安全中文情資相關的文章樣本拓譜圖

KeyGraph 的生成原理，是透過已建立關聯的資料組合進行聚合，將節點之間透過連線建立關係，在網路形成後形成可視化的字詞關聯圖表。

換句話說，當字詞的組合數量過多，表示出現的頻率與事件關聯度越高，越容易形成網路中某個重要節點。

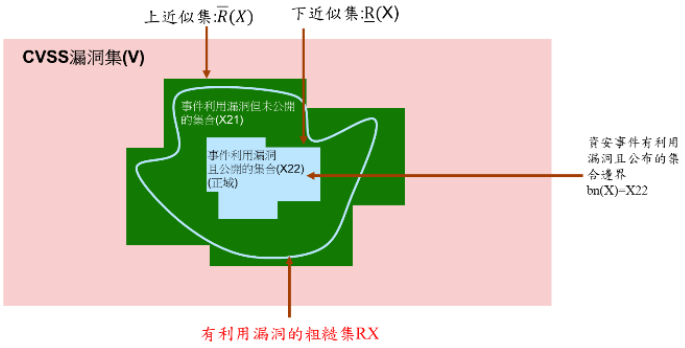
本研究後續將藉由篩選組合的方式，研究主題相關的字詞關聯；例如，以 APT 駭客家族為主題時，強制保留與 APT 有關的組合，並篩選出與該主題相關的組合與關聯，再以深度優先的概念，嘗試將關聯延伸至多個 APT 團體可做關聯，即可分析其相似度及差異性，並進一步探討及研究 APT 家族之間的關聯，檢視 APT 家族間是否存在結構洞現象。

3.3 約略集合理論 (RST, Rough Set Theory)

由於組織在面臨資安事件時，常因各種因素選擇不通報，往往選擇不通報，導致無法掌握真實的漏洞利用率，造成 CVSS 衡量分數的困難。為解決此問題，本研究參考 Al-Daweri, Muataz Salam, et al. (2020) 所提出的方法，採用約略集合理論 (Rough Set Theory, RST) 來計算特徵與類別之間的依賴程度，以推估實際漏洞的利用情況。

本研究進一步透過大量中文情資文本，提取駭客組織與各項攻擊指標，作為駭客家族的特徵進行聚合，並以約略集合理論分析，使用漏洞類別集 (V)，區分為漏洞利用的集合 (U(X)) 與推估的粗糙 RX 集合(有利用漏洞卻未公開的 X21 集合及有利用漏洞

且公開的集合 X22)。



圖三：使用 RST 定出駭客攻擊漏洞行為真實邊界概念

本研究為了掌握 APT 駭客組織利用漏洞的真實特徵，透過下近似集與上近似界集界定出駭客攻擊行為的邊界 (概念詳見圖三) ，並以精確度 (Accuracy) 衡量資訊完整性，最後再與現有的事件情資與漏洞資料庫做比對，結合辨識關鍵攻擊指標 (IoA) 與攻擊手法 (TTPs)，並使用 Fine-tuning 功能微調 Llama 3.1 模型，提升漏洞利用評分之有效性，進而增強資安產品的威脅偵測與預警能力。

3.4 PageRank、TrustRank 相關理論

本研究參考 Liu et al. (2013) 使用 PageRank、TrustRank 和 Anti-Trust Rank 分析反垃圾郵件演算法的方法，將資安事件的攻擊特徵建模為向量 $\vartheta = (\upsilon, \xi)$，其中 $\upsilon = \{e1, e2, \ldots, e_n\}$ 為事件中第 N 個特徵(頂點)，及一個資安事件集合有ξ向量連結的邊[10]。而 PageRank 分數 $r(e)$ 定義為：

$$r(e) = \alpha \cdot \sum_{q \in \in (e)} \frac{r(q)}{|OUT(q)|} + (1-\alpha) \cdot \frac{1}{N}$$

其中，α 是衰減因子，通常設定為 0.85。進一步，可將 Trustrank 理解成反向且有起始點的 Pagerank，而 Pagerank 的分數很高表示其在圖中為高度連結的核心，於社群拓樸中，代表意見領袖，於本研究中則代表 APT 攻擊特徵的中心節點，其影響力會遞迴傳播至整個攻擊網路。

3.5 LoRA 相關理論

[10] 在文本中，對於事件特徵 e，我們使用在(事件特徵)表示連結到它的特徵集合，以及向外(特徵)表示由特徵 e 連結的特徵集合。

研究以 GitHub 上的 CVE-Detail 專案為資料來源，篩選出 2018 至 2025 年 3 月間與 Windows AD 相關的 4,883 筆漏洞樣本，其中 3,818 筆(比例為 78.17%；表一)包含完整的 CVSSv3.x 評分，並依據 Alpaca 資料格式對漏洞描述與八大指標標籤進行結構化組織。接著，採用 LLaMA-3.1-8B 作為基礎模型(演算法詳圖四)，並引入 LoRA

Algorithm 1: Fine-Tuning LLaMA-3.1-8B for Nine CVSS v3.1 Indicator Prediction

Input: Training set $T = \{(d_i, y_i)\}$ where d_i is a vulnerability description and $y_i = (y_i^1, \ldots, y_i^9)$ are the nine CVSS v3.1 indicator labels

Output: Fine-tuned models M_{base}, M_{lora} and their prediction metrics on the CVSS indicators

$M_0 \leftarrow$ `meta-LLaMA-3.1-8B`;
foreach *variant* $\in$ {`Baseline`, `LoRA`} **do**
  if *variant* == `Baseline` **then**
    $M_{\text{base}} \leftarrow$ fine-tune all weights of M_0 on T to minimize cross-entropy over the nine CVSS indicators;
  else
    $A \leftarrow$ attach LoRA adapter to M_0;
    freeze all base weights of M_0;
    $M_{\text{lora}} \leftarrow$ fine-tune adapter A on T to predict the nine CVSS indicators;
  end
  $metrics_{\text{variant}} \leftarrow \text{Evaluate}(M_{\text{variant}}, T)$;
end

圖四：LLaMa 模型透過 Fine-Tuning 做特徵微調演算法

(Low-Rank Adaptation) 技術在分階段學習率與過採樣策略下對模型權重進行高效微調，以增強其對 CVSSv3.x 八大指標之語義映射能力。最後，透過微調前後的多維度性能評估，驗證了本方法在指標識別準確率與動態風險分數生成上的顯著提升與穩定性。

表一：Llama 3.1 模型進行訓練 AD 漏洞樣本數

年度	2021	2022	2023	2024	2025
總筆數	1017	1114	1252	1264	236
有CVSS	708	858	1025	1032	195
無CVSS	309	256	227	232	41

肆、分析與討論 (Analysis and Discussion).

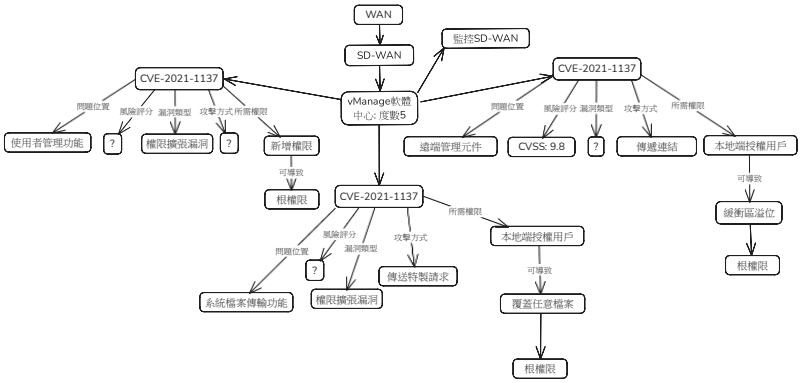
本研究使用拓樸方法，分析 5 年多(2018 年至 2024 年 3 月)台灣論壇所關注的資安事件圖形化，透過節點之間相連，可以更詳細的了解受攻擊者、APT 駭客家族之屬性、漏洞利用，及攻擊態樣間的關係與距離，例如當連接多數節點，且擁有強連結時，此中心點即可能是 APT 駭客家族所使用的重要攻擊手法。

進一步，本研究依照前述研究方法，分析 APT 家族及 CVE 漏洞等情資關聯性變化，並針對 AD 入侵事件的發展趨勢進行探討，得出以下命題及結論:

4.1 高風險(CVSS 評分高)的漏洞，於資安事件群中，具有高中心性等特徵

本研究聚焦於台灣論壇關注之資安事件，將情資文本中關於資訊安全的描述萃取並轉化為具體的技術指標，透過多文本的精煉可歸納出特定駭客組織的行為特徵。以本體 (Ontology) 模組為基底，收集與弱點相關的屬性與指標，建立駭客族群的特徵矩陣． 就此，本研究採用前文斷句法，擷取文本中之情資特徵，並依據 Syed (2020) 的漏洞風險衡量指標，進行人工標記，以建立初步拓譜結構(如下圖思科 SD-WAN vManage 的遠端程式攻擊漏洞情資)，作為後續 Llama 3.1 模型訓練的基礎，計算 AD 高風險 (CVSS 評分高)的漏洞，於拓譜中的各項指標。

研究結果顯示，CVSS 評分較高之漏洞於實際資安事件中的網絡拓樸圖上，顯示出較高的中心性與連結性，詳細驗證資料請參酌命題 2。



圖五：SD-WAN vManage 的遠端程式攻擊漏洞情資拓譜

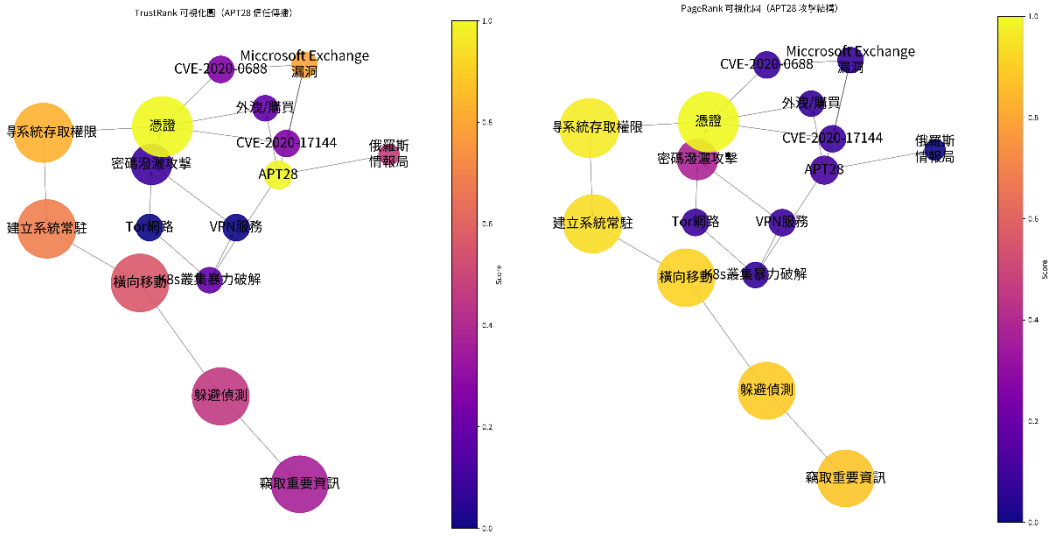
4.2 使用拓樸特徵結合 ATT&CK 框架能有效辨識 AD 的 APT 家族攻擊行為

本研究參考 Li, Jingwen et al. (2024) 研究方法，將 APT 家族視為拓樸圖中的節點，並連結其相關攻擊指標。透過 PageRank 與 TrustRank 等演算法，分析 APT 家族於圖中的連結特徵與行為模式，並進行家族間相似性與關聯性分析。藉由建構特徵矩陣，重新評估漏洞嚴重性，並估算特定族群利用漏洞的可能性。故以 APT28 的 K8s 叢集進行暴力破解攻擊事件作為範例[11]，本研究首先使用 Apriori 演算法將文本進行結構化處理，提取 CVSS 向量指標並對應至 MITRE ATT&CK 的攻擊程序(詳見表一)。再使用拓樸圖建構 APT 行為與攻擊技術間之關聯性，並衡量指標，根據圖五及表二之計算結果，發現憑證竊取與 APT28 節點於 Degree 與中心性指標中皆位居前二。

進一步地，透過 TrustRank 分析(圖六)，以 APT28 節點為信任起點，發現「憑證竊取」與「初始存取」 (例如：導入系統存取權限、建立系統常駐) 節點具有最高信任分數，顯示其為攻擊鏈核心位置；橫向移動與暴力破解則保有中高信任度，反映其在內部擴散階段的重要性；反觀 Tor 網路與 VPN 服務則呈現較低信任分數，顯示其為次要

[11] 例如:文本美英警告：俄羅斯正利用 K8s 叢集進行暴力破解以滲透全球組織，資料來源:https://www.ithome.com.tw/news/145415

支援手段。此結果驗證了 TrustRank 可有效識別具高影響力之攻擊技術節點，並為防禦策略提供具體依據。



圖六：APT28 的 TrustRank、PageRank 相關指標及視化圖

再者，PageRank 分析亦顯示，「憑證」 (Credential) 節點在整體圖譜中得分遠高於其他漏洞與攻擊技術，突顯 APT28 對憑證竊取與濫用的高度依賴，此發現不僅驗證攻擊者的策略重心，也為後續建構多階段攻擊流程與設計針對性防禦措施提供重要參考。

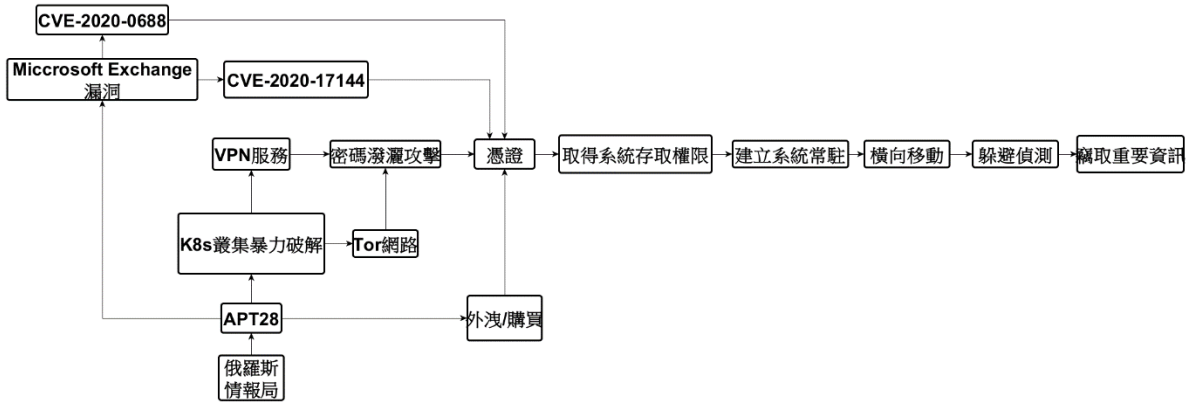
表二：APT28 K8s 暴力破解案例的 MITTER ATT&CK 攻擊程序

Label	Degree	Degree Centrality	Betweenness Centrality	Closeness Centrality	Community
俄羅斯情報局	1.000	0.067	0.000	0.000	0.000
APT28	4.000	0.267	0.067	0.067	0.000
Miccrosoft Exchange 漏洞	3.000	0.200	0.019	0.089	2.000
CVE-2020-0688	2.000	0.133	0.014	0.100	2.000
CVE-2020-17144	2.000	0.133	0.014	0.100	2.000
外洩/購買	2.000	0.133	0.057	0.089	0.000
取得系統存取權限	2.000	0.133	0.210	0.278	3.000
密碼潑灑攻擊	3.000	0.200	0.086	0.152	1.000
K8s叢集暴力破解	3.000	0.200	0.029	0.089	1.000
Tor網路	2.000	0.133	0.021	0.100	1.000
VPN服務	2.000	0.133	0.021	0.100	1.000
建立系統常駐	2.000	0.133	0.171	0.234	3.000
橫向移動	2.000	0.133	0.124	0.209	3.000
躲避偵測	2.000	0.133	0.067	0.192	3.000
竊取重要資訊	1.000	0.067	0.000	0.181	3.000
憑證	5.000	0.333	0.238	0.370	2.000

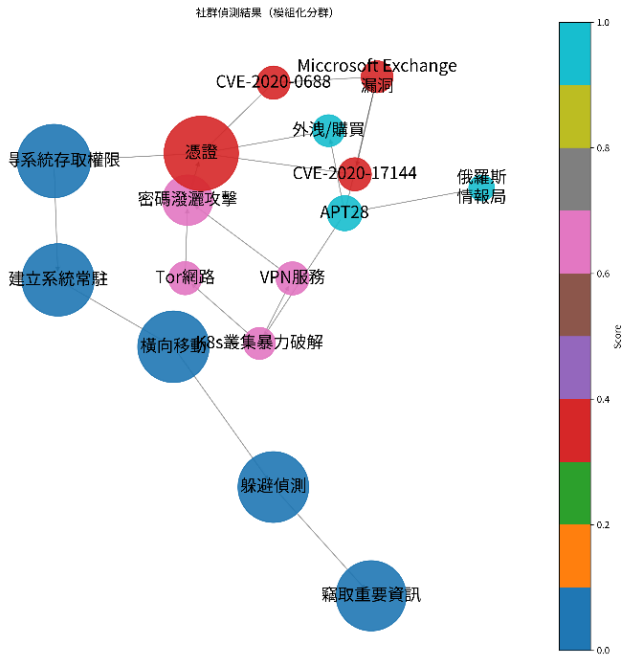
表三：案例 ATT&CK 相關指標

攻擊程序	文中內容
初始存取 (Initial Access)	攻擊者使用外洩或購買的憑證來取得組織的存取權限。 攻擊者開採 Microsoft Exchange 的漏洞（如 CVE-2020-0688 和 CVE-2020-17144）以進入受害組織。
執行 (Execution)	攻擊者利用 Kubernetes 容器叢集執行大規模且分散式的暴力破解攻擊。
防禦規避 (Defense Evasion)	攻擊者透過 Tor 網路或商業 VPN 服務進行攻擊，以避免被追蹤。
憑證存取 (Credential Access)	攻擊者進行大規模的密碼潑灑攻擊，嘗試取得有效的帳號密碼組合。
持續性 (Persistence)	在取得憑證後，攻擊者企圖在受害系統中維持存取權限。
橫向移動 (Lateral Movement)	攻擊者在目標網路中橫向移動，尋找其他可攻擊的系統。
資訊收集 (Collection)	攻擊者竊取各種重要資訊。

此外，關係網絡的時間演變反映了 APT 群體的行為模式隨時間的可能變化。在長時間區間內持續展現相似行為特徵的 APT 群體，於拓樸結構中呈現出更緊密的連結關係，顯示其潛在的協同行動或策略一致性。



圖七：APT28 K8s 暴力破解攻擊事件案例的拓譜



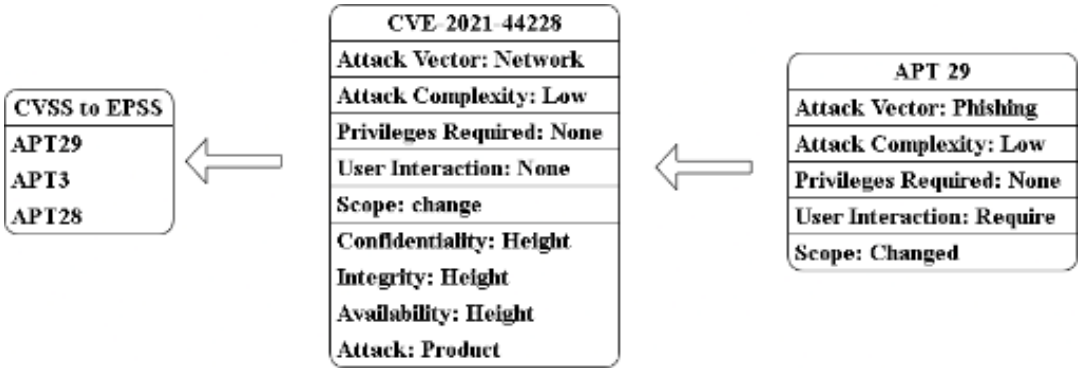
圖八：APT28 的社群拓樸分析及視化圖

Label	Community	PageRank
取得系統存取權限	0	0.124838650167473
建立系統常駐	0	0.121567674023428
橫向移動	0	0.118786882460242
躲避偵測	0	0.116421720934643
竊取重要資訊	0	0.114409466411618
Miccrosoft Exchange 漏洞	1	0.023552595574579
CVE-2020-0688	1	0.0254626448822057
CVE-2020-17144	1	0.0254626448822057
憑證	1	0.128687597180594
密碼潑灑攻擊	2	0.0587392509721581
K8s叢集暴力破解	2	0.023552595574579
Tor網路	2	0.0254626448822057
VPN服務	2	0.0254626448822057
俄羅斯 情報局	3	0.0154528164959305
APT28	3	0.028587575101354
外洩/購買	3	0.023552595574579

圖九：APT28 攻擊節點之社群模組與 PageRank 分分佈圖

4.3 使用 AI 分析能夠建立更有效的風險評分機制

本研究以前述研究為基礎，建立特徵矩陣重新評分漏洞的嚴重性(下圖十)，並計算該漏洞被特定駭客族群使用的機率，計算漏洞受利用及攻擊的風險。



圖十：使用 AI 後的 CVE 風險評分機制

本研究實驗採集 2021 年至 2025 年 3 月與 WindowsAD 漏洞相關樣本統計數據，總樣本數量 4883 筆，有 CVSS 分數樣本為 3,818 筆，比例為 78.17%。初步實驗階段，我們對比了原始 LLaMA 3.1 8B 模型與經 LoRA 微調後模型在 CVSS 八大指標識別任務上的表現差異。結果顯示，LoRA 微調後的模型在各指標的平均識別準確率較微調前提升 超過 12 %(下表四)；其中在 Attack Vector 與 Scope 等核心維度上的精準度增幅最為顯著，驗證了本方法在有限樣本環境中增強模型對漏洞描述與 CVSS 指標語義映射能力的有效性。

表四：CVSS 8 大指標之模型識別準確率與提升幅度比較

指標		A	AC	AV	C	I	PR	S	UI	平均
準確率 (%)	LLaMa3.1	54	9	54	59	51	17	9	46	37.375
	Fine-Tuning	84	86	83	88	85	75	87	86	84.25
提升幅度		30	77	29	29	34	58	78	40	46.875

4.4 使用 AI 分析能夠建立更有效的 ZTA 風險評分機制

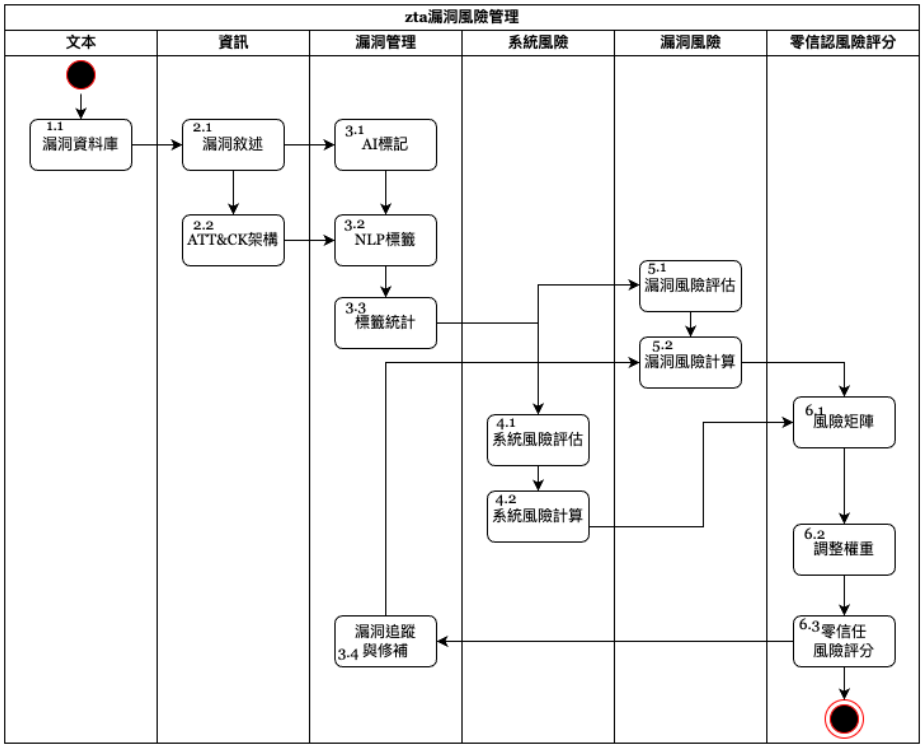
本研究利用前文的風險矩陣評分機制，試圖建構零信任架構 (Zero Trust Architecture, ZTA)[12]之評分機制(詳見下圖十一)，亦即對現有漏洞敘述進行語義標記與關鍵資訊抽取。

[12] 零信任架構 (Zero Trust Architecture, ZTA) 逐漸成為資安防禦的新標準，其核心理念為「永不信任，持續驗證 (Never Trust, Always Verify)」。ZTA 主張不再依賴網路邊界或既定身份作為信任依據，而是針對每一次存取行為，根據主體的設備狀態、認證方式與行為風險進行即時信任判定。這種動態驗證機制，特別能應對企業導入雲端服務、遠端辦公與橫向移動攻擊日益頻繁的趨勢。

透過模型自動分析原始文字內容，萃取出如攻擊方式、認證機制、通訊協定等特徵後，進一步對應至 MITRE ATT&CK 框架中既有的攻擊技術與行為分類。

本研究取 Windows AD 通訊協定相關漏洞為案例，該協定漏洞總數 289 項，在 NVD 漏洞公開資料庫儲存庫 GITHUB 中與未提供 CVSS 分數與 CWE 技術比例分別占 43% 及 47%，使得企業除了難以快速評估該 AD 漏洞外(圖十二；漏洞樣本詳圖十三)，亦讓端點及外部供應鏈產生信任推斷之風險。

就此，本研究所建構的 AI 系統分析顯示，有顯著比例的未評分漏洞與 Active Directory (AD) 相關之攻擊技術高度關聯，尤其在如橫向移動 (Lateral Movement) 占 44%、權限提升 20%及初始化攻擊等戰術層面(表五)，將是 AD 漏洞利用所造成的風險，可作為後續零信任推斷的評分基礎。



圖十一：CVE 漏洞風險建構零信任評分系統

圖十二：Windows AD 通訊協定漏洞未提供 CVSS 分數與 CWE 技術比例

{"cve": "CVE-2025-21242", "description": "Windows Kerberos Information Disclosure Vulnerability"}
[SYSTEM]Windows[/SYSTEM] [AUTH_METHOD]Kerberos[/AUTH_METHOD] [VULNERABILITY_TYPE]Information Disclosure Vulnerability[/VULNERABILITY_TYPE]
{"cve": "CVE-2025-21218", "description": "Windows Kerberos Denial of Service Vulnerability"}
[OS]Windows[/OS] [AUTH_METHOD]Kerberos[/AUTH_METHOD] [VULNERABILITY_TYPE]Denial of Service Vulnerability[/VULNERABILITY_TYPE]
{"cve": "CVE-2025-21217", "description": "Windows NTLM Spoofing Vulnerability"}
[OS]Windows[/OS] [AUTH_METHOD]NTLM[/AUTH_METHOD] [ATTACK_TYPE]Spoofing[/ATTACK_TYPE] Vulnerability
{"cve": "CVE-2024-49127", "description": "Windows Lightweight Directory Access Protocol (LDAP) Remote Code Execution Vulnerability"}
[SYSTEM]Windows[/SYSTEM] [PROTOCOL]Lightweight Directory Access Protocol (LDAP)[/PROTOCOL] Remote Code Execution Vulnerability
{"cve": "CVE-2024-49124", "description": "Lightweight Directory Access Protocol (LDAP) Client Remote Code Execution Vulnerability"}
[PROTOCOL]Lightweight Directory Access Protocol (LDAP)[/PROTOCOL] Client [VULNERABILITY]Remote Code Execution Vulnerability[/VULNERABILITY]

圖十三：標示後的漏洞

表五：依照 CVE 漏洞風險矩陣使用 AI 分辨漏洞分布

漏洞類型	數量	比例
lateral_movement	130	44%
privilege_escalation	59	20%
initial_access	54	18%
credential_access	40	13%
defense_evasion	14	5%

伍、結論 (Conclusion)

本研究以 Windows AD 伺服器攻擊為案例，驗證高風險漏洞在資安情資網絡中呈現的核心樞紐特性，並證實 MITRE ATT&CK 框架結合 PageRank 與 TrustRank 演算法能有效於區隔不同 APT 家族攻擊模式。同時，透過 LLaMA-3.1 模型與 LoRA 微調技術，成功將漏洞描述自動轉換為 CVSS 八大指標，並整合 EPSS 與傳統 CVSS 評分，構建動態風險評分機制。實驗結果顯示，此架構不僅提升漏洞利用預測精確度，也為企業漏洞修補優先順序決策提供科學量化依據。

考量 APT 家族間存在明顯行為模式差異，本研究以特徵拓樸為基礎，成功呈現其攻擊特性與風險鏈結。未來可進一步擴展資料來源，涵蓋多語系、多平台 (如 Linux 與工控系統) APT 情資，以提升模型的跨域適用性與泛化能力。此外，若能結合 SIEM、EDR 平台串流資料並引入時間序列分析，將有助於即時動態更新 APT 家族間的攻擊演化趨勢。在方法層面，建議導入圖像神經網絡 (Graphic Neural Network, GNN) 強化節點表徵學習，並探索新興中心性指標對家族識別與風險預測的增益效益。進一步，透過建構端對端的自動化調校流程，結合測試持續優化效能，可推動 APT 情資分析，朝向即時化、智能化與風險評估導向，作為後續零信任推斷的評分基礎。

參考文獻

[1] Al. Daweri, M. Salam et al. “An Analysis of the KDD99 and UNSW-NB15 Datasets for the Intrusion Detection System,” *Symmetry (Basel)* 12.10 (2020): 1666-. Web.

[2] A. Rezapour, and W. G. Tzeng. “A Robust algorithm for predicting attacks using collaborative security logs,” *Journal of Information Science and Engineering* 36.3 (2020): 597–619. Web.

[3] Basheer, Randa, B. Alkhatib, and Z. Xu. “Threats from the Dark: A Review over Dark Web Investigation Research for Cyber Threat Intelligence,” J*ournal of Computer Networks and Communications* 2021 (2021): 1–21. Web.

[4] Chng, Samuel et al. “Hacker types, motivations and strategies: A comprehensive framework,” *Computers in Human Behavior Reports 5* (2022): 100167-. Web.

[5] de Martí Beltran, Joan. “Matthew O. Jackson, Social and Economic Networks, Princeton University Press (2008),” *Regional Science and Urban Economics 39.5* (2009): 644–645. Print.

[6] Jacobs, Jay et al. “Exploit Prediction Scoring System (EPSS),” *Digital threats* (Print) 2.3 (2021): 1–17. Web.

[7] J. Vukalovic, and D. Delija. “Advanced Persistent Threats - Detection and Defense,” *2015 38th International Convention on Information and Communication Technology*, Electronics and Microelectronics (MIPRO). MIPRO, 2015. 1324–1330. Web.

[8] Koloveas, Paris et al. “InTIME: A machine learning-based framework for gathering and leveraging web data to cyber-threat intelligence,” *Electronics (Basel)* 10.7 (2021): 818-. Web.

[9] Lamprakis, Pavlos et al. “Unsupervised Detection of APT C&C Channels Using Web Request Graphs,” Detection of Intrusions and Malware, and Vulnerability Assessment. vol. 10327. Switzerland: Springer International Publishing AG, 2017. 366–387. Web.

[10] Lee, J. San et al. “ML-Based Intrusion Detection System for Precise APT Cyber-Clustering,” *Computers & Security 149* (2025): 104209-. Web.

[11] L. Robert, H. Janicke, and S. Schrittwieser. “AIDIS: detecting and classifying anomalous behavior in ubiquitous kernel processes,” *Computers & Security 84* (2019): 120–147. Web.

[12] Li. Jingwen, J. Liu, and R. Zhang. “Advanced Persistent Threat Group Correlation Analysis via Attack Behavior Patterns and Rough Sets,” *Electronics (Basel)* 13.6 (2024): 1106-. Web.

[13] Lin, P. Ching et al. “Correlation of cyber threat intelligence with sightings for intelligence assessment and augmentation,” *Computer Networks* (Amsterdam, Netherlands : 1999) 228 (2023): 109736-. Web.

[14] Liu, Xinyue et al. “Combating web spam through trust–distrust propagation with confidence,” *Pattern Recognition Letters 34.13* (2013): 1462–1469. Web.

[15] M. Liras, L. Francisco, A. R. d. Soto, and M. A. Prada. “Feature analysis for data-driven apt-related malware discrimination,” *Computers & Security 104* (2021): 102202-. Web.

[16] Ussath, Martin, et al. “Advanced persistent threats: Behind the scenes,” *2016 Annual Conference on Information Science and Systems (CISS)*, IEEE, 2016, pp. 181–86, https://doi.org/10.1109/CISS.2016.7460498.

[17] R. Amit, T. Berenblum, and D. Maimon. “The secondary global market for hacked data,” *International Journal of Cyber Criminology 12.2* (2018): 408–426. Web.

[18] Syed, Romilla. “Cybersecurity vulnerability management: A conceptual ontology and cyber intelligence alert system,” *Information & Management 57.6* (2020): 103334-. Web.

[19] Tan, Cheng et al. “Attack provenance tracing in cyberspace: Solutions, challenges and future directions,” *IEEE network 33.2* (2019): 174–180. Web.

[20] W. T. En, L. J. Kai, H. T. Chen, C. C. Hai, L. S. Ying, W. J. Fu , “STIE-ZTA: Self-adaptive trust inference engine for policy-based zero trust architecture,” *Cryptology and Information Security Conference 2025*.

[21] 王仁甫、陳品翰、陳思宥、呂書晴、林瑜君、魏得恩 (2025)，AI 方法建構 APT 家族攻擊情資特徵拓樸及風險矩陣- 以 AD 伺服器攻擊為案例研究，第三十五屆全國資訊安全會議