工控資安發展新思維 - TA-DIF

TA-DIF 打造可解釋可部署 AI OT 安全異常偵測機制

OT 資安逐漸受到重視，針對快慢攻擊的 AI OT 安全架構設計的 TA-DIF，即是要打造一個可解釋、可部署的異常偵測機制，如此，在面對主管機關政策推進與資安挑戰深化情勢下，TA-DIF 提供工控資安管理方面的 AI 新思維。

文／梁日誠

在數位轉型與智慧化浪潮推動下，OT（Operational Technology）與 ICS（Industrial Control Systems）正面臨前所未有的資安壓力，傳統部署於 IT 環境的資安設備，如 IDS（入侵偵測系統）與 IPS（入侵防禦系統），在 IT 環境能發揮預期作用，但在 OT 環境中卻顯得力有未逮，特別是在面對慢速滲透攻擊（Slow Attack）與非典型行為操控（如：資料偽造、程序錯置）時，偵測精準度與即時性大幅下降。

更嚴峻的是，在零時差攻擊（Zero-Day Attacks）與目標導向型異常（如：APT 進階持續性威脅，OT-Layer Attack）日益頻繁的現實狀況下，傳統 Signature-Based 防禦體系往往來不及反應。

而現今企業若想進一步導入 AI 進行資安分析，AI 模型準確度再高，若無法解釋系統何以發生異常，甚至無法分析其根本原因，如此將會失去第一線的 OT 人員對系統運作的信任。

OT 安全除了影響可用性（Availability）、完整性（Integrity）、機密性（Confidentiality）外，也對健康（Health）、安全（Safety）、環境（Environment）造成衝擊，更與關鍵基礎設施所涉及的國家安全（National Security）息息相關，唯有化被動為主動，才能根本的解決痛點。

工控資安亟需新一代偵測架構

推動 OT 資安的障礙，部分來自於人才結構與資源現況的限制，一方面 OT 現場人員熟稔製程邏輯與控制作業，但對於 AI 模型、資安策略往往陌生，這也讓企業高層在導入先進資安架構時進退維谷。

另一方面，許多已運行多年的 OT 場域，其系統與設備高齡化、模組封閉、在地與原廠支援皆捉襟見肘，先不考慮資安優化，某些場域在基本維運都顯得吃力，更遑論導入 AI 式的安全防禦機制。

在這樣的背景下，也就迫切需要一種不依賴大量攻擊資料、不仰賴雲端運算、可在地端落地部署、同時兼顧可解釋性與攻擊適應性的異常偵測機制，而這正是「**TA-DIF（Time-Adaptive Dual Isolation Forests）**」適用的效益。

值得注意的是，OT 資安事件已不再只是少見的情境，以近年美國佛羅里達州 Oldsmar 市的淨水處理廠即曾遭駭客入侵為例，攻擊者透過桌面共享軟體 TeamViewer 遠端控制 SCADA 系統，意圖將水中氫氧化鈉濃度升高至危險值，所幸現場人員及時發現並回復設定值，避免一場災難。

但事後調查揭露多項弱點，包括，使用已停產支援的 Windows 7、未設防火牆、使用共用帳密、外部連網未隔離、缺少即時有效的異常偵測系統，凸顯基層 OT 現場資安防護的實質斷層，這類事件

也再次強化了「**偵測即是防禦**」的重要性，尤其在設備更新困難、可用資源有限的環境中，更需部署能夠因地制宜，具備前置性（Proactive），且能即時警示的 AI 式異常偵測架構。

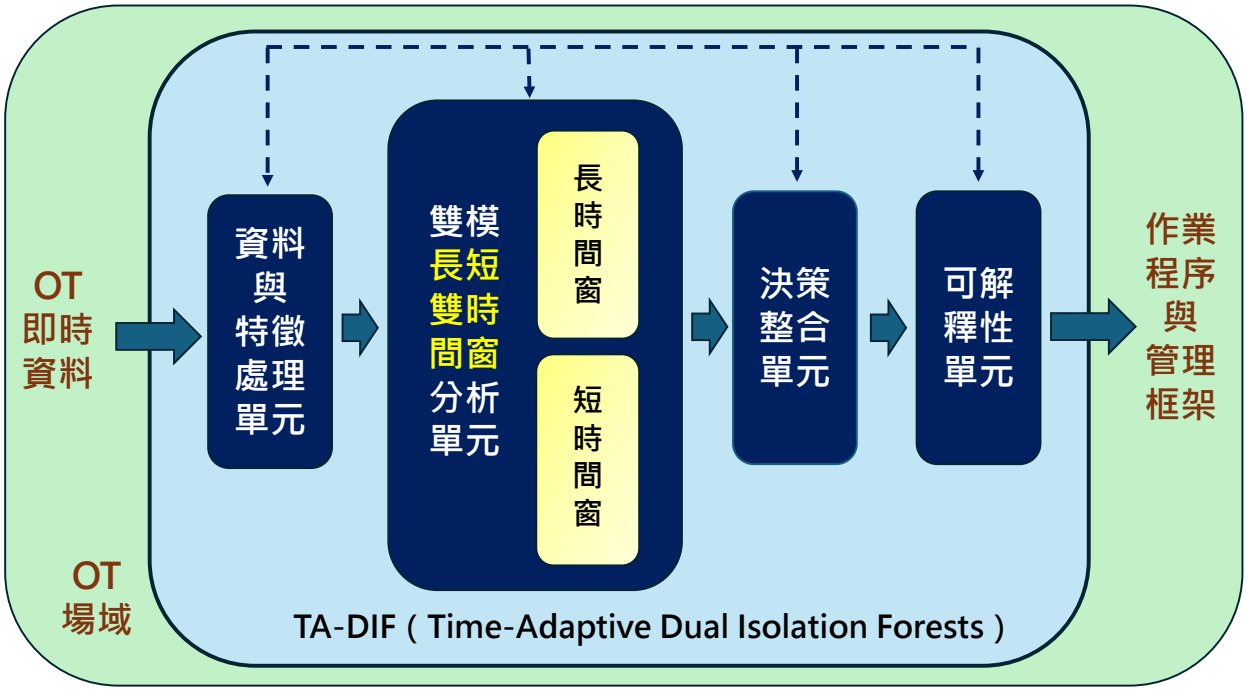
TA-DIF 同時兼顧快攻與慢攻的偵測架構

TA-DIF 架構是建立在 Dual Isolation Forests（雙模型）的機器學習 AI 基礎之上，分別以**原始感測特徵與主成分分析（Principal Component Analysis；PCA）降維特徵**訓練兩個 Isolation Forest（IF）模型。其進一步優化之處，在於導入兩個時間窗（Window，w）設計，一為時間短窗（如：120 秒）來對應快攻（Burst Attack 爆發攻擊）；一為時間長窗（如：600 秒）對應慢攻（Slow Attack 緩慢攻擊）。

只要任一模型於任一時間窗偵測出異常，系統即可在本地立即發出告警，無需連網或依賴雲端判斷，此策略大幅強化了對持續性緩慢攻擊（如：數據滲透、操控延遲、假穩定值等）的敏感度，同時保留對 Burst-Type 快速異常或攻擊的反應速度。

在實務運作上，OT 即時資料流（如：感測器值、致動器狀態等）會同步導入 TA-DIF 架構（如圖）中的兩組模型：IF-1（如：w = 120 sec）與 IF-2（如：w = 600 sec），彼此以不同的時間尺度分析同樣的資料來源，這種資料同步處理的方式，讓系統在偵測異常時具備時間上的「雙焦點」，兼顧即時性與趨勢偏移的可見性。

在 AI 可解釋性的議題上，TA-DIF 的兩個模型亦具備回溯異常來源的能力。IF-1 採用原始感測特徵，可透過統計各棵樹中分裂節點的特徵出現頻率，而推論哪些感測器或致動器的變數貢獻最大，進而指出異常可能來自哪個實體元件。至於 IF-2，則基於 PCA 降維後的主成分，可透過 PCA Loading Matrix，將異常表現突出的主成分反推至原始變數的加權貢獻，從而定位具高風險變異的來源感測器或控制參數。



從架構分析，TA-DIF 包含以下主要單元：

1. 資料與特徵處理單元：執行資料預處理、特徵工程與降維。
2. 雙模型單元：IF-1 使用原始感測特徵資料，IF-2 使用 PCA 降維後的主要特徵子空間。
   如架構圖中標示「長時間窗」與「短時間窗」處。
3. 雙模長短雙時間窗分析單元：分別以「長時間窗」與「短時間窗」，進行異常偵測。
4. 決策整合單元：若任一模型於任一時間窗期間判為異常樣本比例超過門檻（如：80%），即觸發告警。
5. 可解釋性單元：追蹤 Isolation Score、對應主要分裂特徵，提供異常元件或變數建議。

TA-DIF 這種雙向可追蹤能力，為 OT 現場工程師提供了理解與應變依據，且在實務中足以建立現場人員信任的關鍵因素之一。

此外，TA-DIF 並可與標準作業程序（如：異常處理程序）與資安監控程序進行整合，來支援完整的事前偵測預防、事中防護處理與事後回復改善的資安管理框架與制度。

TA-DIF 讓資安 AI 可落地應用

與深度學習或需要大量樣本訓練的監督式學習模型不同，TA-DIF 架構採用非監督式學習技術，只需正常的資料流資料即可建立模型，降低了 OT 場域的資料準備與維運壓力，整個架構可在 Windows 11 電腦（如：Log Server、工控電腦）上離線執行，不需 GPU 或 TPU、不依賴雲端。

可用 Python 執行並分析 CSV 或 Log 檔，模型可封裝為「.pkl」，部署維運簡便，特別適合於電力、水、石油、製造、交通、醫療、國防等有 Air-Gap 限制的（類）工控網路環境。

此外，TA-DIF 可進一步與現場標準作業程序整合，將異常樣本對應至控制元件（如：閥門、泵浦、感測器）進行根本原因對照分析，有助於第一線技術人員判讀與應變，這也彌補了 AI 資安模型在 OT 環境中「技術可信度不易建立」的障礙。

公司管理階層與資安長在面對 OT 資安管理與治理時，經常在現實限制與資安期待之間拉鋸，TA-DIF 提供一個兼具技術可行性、政策適應性與組織落地彈性的切入點，不需全面性改造設備與網路，即可在既有 OT 資料流中建立異常偵測防禦機制。

資安並不是一次性的產品或服務的部署，亦包含持續性的資料理解與決策輔助，TA-DIF 是協助企業邁出這一步的起點，也是一個值得列入 OT 資安藍圖的技術選項。

從 AI 與資安治理的雙重合規角度而言，於 OT／ICS 環境中建立與 IT 資安標準（如：ISO 27001）互補的 IEC 62443-2-1 國際標準管理架構，進而與 ISO 42001 人工智慧管理整合，是企業展現 OT 資安合規與韌性治理的有效方式。

而 TA-DIF 架構的實作，更體現出 Proactive（主動預防）與 Interactive（具現場互動解釋）兼備的資安作為，恰補足現場反應與系統監控間的最後一哩路。

政府資安政策聚焦 OT 資安

數位發展部資安署於日前新聞指出，未來施政將在既有基礎上，透過公私協力與各方積極對話，所聚焦四項重點工作中，包含「提升關鍵基礎設施資安防護及韌性」，確保臺灣資安防線的堅韌，打造可信賴的資安環境；並表示於「提升關鍵基礎設施資安防護及韌性」方面，工控系統安全將是重要挑戰。

將透過建立資安防衛體系，加強工控系統人員培訓，增加關鍵基礎設施構面專家，落實領域基準以提升關鍵基礎設施工控系統防護能量。

TA-DIF 的可部署性、可解釋性與模型輕量化特性，正好成為主管機關政策與實務需求之間的技術橋梁，也有助於形成可擴展的資安治理與管理架構，協助企業邁向自主管控與風險可視的下一階段。

作者：梁日誠 Lead CCAI CCPI PII CISSPI CISSOI CPTEI CCSOI CDREI CCISOI ISMI ISAI AIMPI FIAAISI FHCA- EU AI ActI CAIEI CAIPCI CDEI CDAI DSFMI FHCA-GDPRI IDPPI ICEPI GRCAI GPM-bI IMPCI，Pending，現為加拿大 SCC／MC ISO／IEC JTC1／SC42、SC27、ISO／TC22／SC32、IEC／TC65 技術組成員，ISO 42001／ISO 27001／ISO 27701／ISO 22301／ISO 20000-1／IEC 62443-2-1 稽核師及講師，TCIC 環奧國際驗證公司全球營運總經理。