零信任與精準 AI 可加強釣魚郵件防禦

資安威脅更將無所不在（3）

釣魚郵件雖然是顯而易見的資安威脅，過去可以透過社交工程加強員工資安意識來降低誤觸造成的資安缺口，然而，在 AI 的介入後，釣魚郵件將會超進化，多數人將難以辨別而升高資安上的威脅程度。

採訪／施鑫澤　文／林裕洋

蔓延超過三年的 COVID-19 疫情，不光改變全球經濟結構與人類行為，也讓網路釣魚攻擊速度、複雜度等急速增加。由於在家工作環境的壓力、焦慮和持續挑戰，也讓駭客開始針對以企業員工為對象，發送表面上看起來真實的內容，引誘使用者點擊其實不應該點擊的內容，進而達到入侵的目的。

儘管隨著勒索軟體攻擊盛行，根據 Palo Alto Networks Unit 42 公布研究報告指出，2022 年網路釣魚攻擊比例約達 33%，然到 2023 年下降到僅剩 17%。不過隨著 AI 熱度不斷攀升，未來 6～12 個月主流手法是利用 AI 強化社交工程攻擊，例如網路釣魚和商業電子郵件詐騙（BEC）等快速攀升。這正突顯出網路釣魚在 AI 的加持下，再度成為重要威脅之一。

Palo Alto Networks 台灣區技術總監蕭松瀛指出，網路釣魚攻擊並不是新問題，也非疫情爆發時才開始發生，近來成為主流的網路釣魚套件工具，為攻擊者提供有效的網路釣魚即服務，乃至於包含能規避偵測的現成功能。正因如此，時至今日，企業仍然遭受網路釣魚攻擊威脅，成為資安團隊必須克服的挑戰。駭客藉由 AI 語言模型協助，能研究攻擊目標的完整電子郵件記錄和通訊模式，進而製作毫無破綻、以假亂真的網路釣魚訊息。這類自動化網路釣魚威脅已開始發生，且影響規模也持續擴大中，預期將會在短期內更加普遍，成為攻擊者首選的攻擊方法。

AI 技術加持　攻擊手法大幅進化

隨著 AI 技術持續進化，預估 2030 年前 AI 可望在所有產業的效率、體驗和成長方面掀起一波新浪潮。根據勤業眾信研究報告指出，2032 年 AI 網路安全市場預計將達到 1,027.8 億美元。儘管 AI 為企業帶來優勢，但因駭客也大量利用 AI 技術發展新型態攻擊受法，如進行傳統安全解決方案無法偵測到的規避、獨特和破壞性的零日攻擊、網路釣魚攻擊，成為資安團隊面臨的新挑戰。

事實上，新一代網路釣魚攻擊手法都相當地隱蔽，不會使用惡意軟體避免觸發常見部署的偵測技術，而是採用許多不同混淆技術來規避傳統的 Web 安全掃描，常為駭客能突破網路安全設備偵測的重要關鍵。

蕭松瀛表示，駭客選擇運用 AI 語言模型與網路釣魚結合關鍵，是發現人們會在電子郵件往返中閒話家常，如談及狗狗或家人的近況，而 AI 語言模型則具備複製這種對話模式。AI 語言模型在培養足夠的信任後話鋒一轉，提出「對了，可以填一下那張發票嗎？」，就像 6 個月前填過的那樣。此時受害者因警戒心降低，多半願意配合、

Palo Alto Networks 台灣區技術總監蕭松瀛（Nicholas）。

更容易受騙。

Palo Alto Networks 認為未來 12 個月到數年之間，駭客會從兩個關鍵領域著手，進一步運用 AI 發展攻擊能力。首先是製作惡意軟體，以現有惡意軟體程式碼訓練 AI 語言模型，並組合成全新病毒變種，藉此躲過資安設備偵測。其次，則是針對 AI／ML系統進行攻擊，預計會使用提示注入（prompt injection）和訓練資料集植毒（poisoning training dataset）等技術，藉此操縱模型的輸出結果。由於愈來愈多企業運用大語言模型結合生成式 AI 技術，未來會有駭客嘗試對訓練資料下毒與發動攻擊。

蕭松瀛指出，預計五年後 AI 將可同時對數千到數百萬個目標發動自主攻擊，而能大幅擴張規模，遠遠超出人類迄今所做過的攻擊活動。駭客在 AI 助力下，能以自動化方式判定攻擊目標、橫向移動，以及擷取有價值的資料。屆時，企業將面臨完全自動化的入侵對手，因此也必須防範此類採用全自動攻擊手法的駭客團體。

以 2020 年 SolarWinds 供應鏈攻擊事件為例，當時受影響企業與政府單位超過上萬家以上。儘管當時惡意軟體廣泛散播，但仍只能利用一小部分後門進行攻擊，關鍵在於駭客團隊的資源和人力依然有限。然而若此類攻擊手法融合 AI 之後，可望以自動化方式完成複雜攻擊工作，屆時駭客將不會攻擊數百個網路，而是可能高達數千個網路，且還會植入誘餌，等待日後時間一到發動攻擊。

運用 AI 強化偵測力　揪出惡意威脅

雖然 AI 成為駭客發展新型態惡意軟體的工具，但資安公司也開始運用此技術強化網路安全防禦能力。在 AI／ML 技術下進步下，軟體開發過程中可透過自動化測試和大規模修復的能力，減少可能被駭客利用的漏洞數。

此外，近期幾乎成為資安顯學的「零信任」，不管是觀念、機制，或是架構，都可大幅降低資安上的風險，然而，知易行難，要想打造出密不透風的零信任機制，非一蹴可幾且相關預算費用也較難以掌控。

因此，在資安營運方面，資安團隊可透過 AI／ML 技術協助，專注於提升主動防禦技術，例如威脅獵捕（threat hunting）而非被動的漏洞分檢（triage）。因為若能自動化處理 SOC 已知所有威脅，SOC 就能轉向關注威脅獵捕等主動防禦任務，有助於大幅提升創造力，並發現未知的新型攻擊。

「資安團隊工作效率高低，在於能有多少時間用於搜尋新威脅，所以若能妥善運用 AI 技術協助，有助於降低營運團隊的成本支出，減少團隊處理已知攻擊或誤報的事件，而是能建立一套防堵未知威脅的策略。」蕭松瀛解釋：「隨著 AI 網路攻擊和防禦能力同步進展，資安團隊若能運用 AI 潛在優勢，可望清楚掌握新興威脅態勢，妥善保護公司重要資產。」

創新精準AI　抵禦駭客多元手法

根據 Palo Alto Networks Unit 42 公布 2024 勒索軟體威脅報告指出，2023 年全球多重勒索軟體攻擊增加 49%，在台灣市場部分，製造業、高科技業成為受攻擊最猛烈的重災區。此關鍵在於 AI 技術提供駭客更多自動化攻擊手段，導致網路攻擊規模持續擴大，攻擊速度從過往數天之內的攻擊早已快到變成數小時，成為資安防護團隊的夢魘。

為此，該公司推出一系列全新安全解決方案，幫助企業阻擋 AI 生成式攻擊，並有效保

護 AI 設計的安全。該公司運用創新精準 AI（Precision AI）結合機器學習和深度學習的最佳效能與即時生成式 AI 存取能力，實現更主動的網路和基礎設施保護措施，採取憑藉 AI 驅動的安全性先發制人機制。

蕭松瀛表示，面對駭客透過 AI 發動前所未有的快速、自動化攻擊趨勢，企業亦必須運用 AI 技術進行對抗，才能快速應變、避免龐大經濟損失。Palo Alto Networks 精準 AI 技術將帶來顛覆以往的改變，創造前所未有的新安全典範，此技術能在短時間內分析大量的資料，加快偵測速度和準確性，從而大幅縮短平均回應時間（MTTR）和平均偵測時間（MTTD）。簡單來說，我們將過往需耗費 2～3 天以上的回應時間，縮短到數小時、數十分鐘內完成，讓企業能在第一間回應威脅。

Palo Alto Networks 平均每天可以發現 230 萬個前一日不存在的新型獨特威脅，平均每天阻止 113 億個內嵌威脅，這正代表實現平台化和採用精準 AI 的優勢。

運用平台化概念　降低設備維運難度

前面提到，現今企業逐漸感受到能夠整合跨網路、雲端和資安監控中心（SOC）環境安全平台所帶來的多種優勢，如避免發生不同解決方案之間的隔閡，還能簡化打造資安架構的複雜度，進而與提高營運效率。

精準 AI 最大特色，是將該技術整合進 Strata、Prisma 與 Cortex 等產品之中，運用平台化概念整合功能和資料可存取性，實現運用 AI 對抗對抗 AI 概念。該公司推出的精準 AI 安全套組，訴求由精準 AI 提供支援的進階安全服務，包括進階網址過濾、進階威脅防禦、進階防火牆和進階 DNS 安全性。這些服務運用內嵌 AI 來防止複雜的網路型威脅、零日威脅、規避性命令與控制攻擊，以及 DNS 綁架攻擊。

蕭松瀛指出，該公司非常瞭解資安團隊運用多套資安設備打造資安防護平台時的痛點，以及後續維運上的挑戰。我們以精準 AI 概念提供的網路安全平台，內建多種 AI 自動化工具，就是希望讓企業透過自動化機制對抗駭客的自動化攻擊手法。我們不光提供平台解決方案，也能提供企業所需要的資安顧問服務，讓資安人手不足的企業，也能解決各種資安挑戰與困境。

精準 AI保護 AI 模型安全

Palo Alto Networks 推出精準 AI 主要建立在三大基礎上，首先「用 AI 對抗 AI」，訴求面對駭客運用 AI 技術攻擊企業網路和資料之前，運用精準 AI 支援的 AI 安全性進行即時防禦，進而達到偵測並防止多型態威脅的加速。其次，則是透過設計確保 AI 安全，前面提到駭客開始攻擊 AI 模型中的資料，精準 AI 可保護組織的 GenAI、AI 支援的應用程式開發，讓資安團隊享有基於 AI 的可視性、控制和保護，包括防止提示植入攻擊。最後，則是運用AI簡化網路安全、降低複雜度，並且提供可採取行動的見解，透過逐步提出建議來引導資安團隊解決問題。

蕭松瀛說，我們在 AI 執行階段安全性，則是透過保護整體 AI 應用程式生態系統，防止執行階段威脅，如提示注入、模型 DoS、不安全輸出等，助企業建立 AI 驅動的應用程式。此外，我們也提供全新支援 AI 的程式碼到雲端（Code to Cloud）功能，包括 AI 攻擊路徑和影響範圍分析、揭示複雜風險的引導式補救措施、從初始攻擊擴散風險的潛在漏洞途徑，以及快速進行補救的步驟。