Comando e controllo su un broker MQTT pubblico: le backdoor Toy Ghouls e il punto cieco delle reti industriali

Due impianti Windows analizzati da Kaspersky non usano infrastruttura propria per il comando e controllo: uno si appoggia al broker MQTT pubblico di HiveMQ, l'altro al protocollo Matrix. Le vittime documentate sono russe e non si tratta di un attacco a sistemi industriali, ma la tecnica solleva un problema di rilevamento che riguarda direttamente chi gestisce reti OT.

Due backdoor, nessuna infrastruttura propria

Il 4 settembre 2026 Kaspersky ha pubblicato su Securelist l'analisi di due backdoor che i suoi ricercatori attribuiscono al gruppo a movente finanziario noto come Toy Ghouls, indicato anche con i nomi Bearlyfy, Laboo.boo e Feral Wolf. L'attribuzione è dei team GERT e Security Services di Kaspersky e va riportata come tale.

Il gruppo è attivo dal 2025 contro organizzazioni russe. Ha iniziato appoggiandosi a strumenti disponibili pubblicamente e a *builder* di ransomware trapelati, quelli di Babuk e LockBit, per poi sviluppare codice proprio, tra cui il ransomware GenieLocker, in versioni per Windows, Linux ed ESXi.

I due impianti individuati all'inizio di luglio 2026 si chiamano mqtt-bird-agent 0.1.0 e matrix-bird-agent 0.1.0, e corrispondono a due eseguibili Windows, cplsupport.exe e

wtass.exe. Il tratto che li rende interessanti è che nessuno dei due porta con sé un server di comando e controllo: entrambi si appoggiano a servizi e protocolli legittimi.

Un secondo elemento merita attenzione, e riguarda il modo in cui arrivano sulla macchina. La distribuzione avviene attraverso Windows Remote Management, con utilità di post-sfruttamento *open source* come Evil-WinRM e WinRM-fs. Il report non descrive il vettore di accesso iniziale, ma la scelta di WinRM implica che gli operatori dispongono già di credenziali amministrative quando installano l'impianto: la backdoor non è il punto di ingresso, è ciò che rende l'accesso persistente.

La persistenza si ottiene registrandosi come servizio Windows, con i nomi "cplsupport" (descritto come "Problem Reports Control Panel") e "wtas" ("Windows Telemetry Aggregator Service"), tramite argomenti di installazione e disinstallazione. In entrambe le varianti i campi sensibili della configurazione vengono cifrati al primo avvio con ChaCha20-Poly1305, con una chiave derivata dal valore MachineGuid del registro di sistema: un accorgimento che impedisce di riutilizzare la configurazione su un'altra macchina e complica il lavoro di chi analizza il campione fuori dall'ambiente di origine. La variante MQTT conserva il file cifrato su disco; la variante Matrix lo cancella dopo il primo avvio e sposta i parametri in una chiave di registro dedicata.

Perché un broker MQTT pubblico è un canale quasi ideale

Secondo Kaspersky, la variante mqtt-bird-agent comunica con broker.hivemq.com sulla porta 8883, quella standard di MQTT su TLS. Il dominio è quello del broker pubblico di HiveMQ, uno dei più noti tra quelli messi a disposizione gratuitamente per test e prototipazione: non è un server compromesso né un dominio registrato dagli attaccanti. Il report descrive

quattro percorsi di scambio: uno per lo stato del sistema infetto, uno per le metriche periodiche (CPU, memoria, disco, carico, tempo di attività), uno da cui la backdoor legge i comandi e uno su cui restituisce i risultati. I comandi vengono eseguiti tramite PowerShell in modalità nascosta.

Un'annotazione di precisione: nello stesso passaggio Securelist parla di un "cluster" creato dagli attaccanti e di un piano gratuito con limiti di 100 connessioni concorrenti e 10 GB di traffico mensili. Sono esattamente le caratteristiche del piano Serverless gratuito di HiveMQ Cloud, che prevede credenziali di accesso e TLS, non quelle del broker demo aperto, che non ha né cluster né piani. Il report non scioglie l'ambiguità. La conseguenza pratica è che non si può dare per certo che il canale fosse privo di autenticazione, e che l'indirizzo effettivamente contattato potrebbe essere un sottodominio dedicato del servizio cloud piuttosto che il nome generico del broker pubblico: in entrambi i casi, comunque, la destinazione appartiene a un fornitore legittimo e ampiamente usato.

Vale la pena spiegare perché questa scelta sia efficace, al di là della gratuità.

MQTT è un protocollo *publish and subscribe*: il client non riceve connessioni in ingresso, si limita a stabilire una connessione uscente verso il broker e ad attendere. Non serve alcuna porta aperta sul perimetro, e il traffico ha la forma di una sessione TLS lunga verso un servizio cloud, esattamente come mille integrazioni legittime. I messaggi sono minuscoli, perché il protocollo nasce per la telemetria su reti a banda ridotta, quindi il volume non fa scattare soglie. E poiché il broker è condiviso da chiunque lo usi per test, l'indirizzo di destinazione non è un indicatore di compromissione: bloccarlo a livello di reputazione significa bloccare un servizio legittimo.

C'è però un motivo ulteriore, ed è quello che riguarda

direttamente il nostro perimetro editoriale. MQTT è lo standard di fatto della telemetria industriale e dell'*Internet of Things*. In una rete di stabilimento, in un impianto di trattamento acque, in una flotta di dispositivi connessi, il traffico MQTT in uscita non è un'anomalia: è il funzionamento normale. Un canale di comando e controllo che parla MQTT, in quel contesto, non si nasconde tra il rumore, coincide con il rumore.

Il caso Matrix, e la differenza che quasi nessuno sta facendo

La seconda variante usa il protocollo Matrix. L'agente invia messaggi di stato e metriche con tipi di evento personalizzati, riceve comandi preceduti dal prefisso cmd: da un account chiamato "panel-bot" e li esegue attraverso la riga di comando di Windows, non PowerShell. Accetta inoltre un'istruzione di configurazione che regola l'intervallo di invio delle metriche, in una finestra che va da cinque secondi a un'ora; il valore viene salvato nel registro. Le risposte tornano come eventi di un tipo dedicato.

Qui però la somiglianza con il caso MQTT si ferma, e conviene dirlo chiaramente perché diverse riprese della notizia stanno appiattendo le due cose. Il server Matrix utilizzato è sotto il controllo degli attaccanti, raggiungibile a un dominio che richiama il nome del client Element. Non c'è quindi abuso di un servizio di terzi: c'è l'uso di un protocollo legittimo e federato su infrastruttura propria. È mimetismo di protocollo, non parassitismo di servizio.

La distinzione conta per chi deve difendersi. Nel caso MQTT il dominio di destinazione è inutilizzabile come indicatore, perché legittimo; nel caso Matrix il dominio è un indicatore valido e bloccabile, ed è l'analisi comportamentale a essere più difficile, perché il traffico Matrix cifrato è opaco per definizione.

Un’osservazione di prospettiva, che è nostra e non del report. Matrix sta guadagnando terreno nelle amministrazioni pubbliche europee proprio come alternativa sovrana alle piattaforme di messaggistica statunitensi, e il tema della sovranità digitale ne sta accelerando l’adozione. Nelle organizzazioni che lo hanno introdotto, il traffico Matrix passerà presto dall’essere un’anomalia all’essere ordinario. Vale per Matrix ciò che vale già oggi per MQTT negli impianti industriali: la protezione offerta dall’anomalia statistica ha una scadenza, e coincide con il momento in cui la tecnologia diventa diffusa.

Che cosa questo caso non dice

Poiché l’angolo più interessante riguarda le reti industriali, è necessario essere precisi su ciò che il report documenta e su ciò che non documenta, altrimenti l’analisi si trasforma in allarmismo.

Non si tratta di un attacco a sistemi di controllo industriale. Gli impianti girano su Windows, vengono installati come servizi dopo che gli attaccanti hanno già ottenuto privilegi amministrativi, ed eseguono comandi tramite PowerShell o riga di comando: è una backdoor IT, non un *malware* progettato per interagire con logiche di processo. Le vittime documentate sono organizzazioni russe, e il report non individua settori specifici né fornisce un numero di vittime.

Il motivo di interesse non è quindi ciò che è accaduto, è ciò che la tecnica rende possibile. Un canale di comando e controllo su MQTT esiste ora, è documentato e funziona. Il giorno in cui un impianto simile venisse installato su un sistema di una rete dove MQTT è traffico normale, la difficoltà di rilevamento non sarebbe teorica. È un’osservazione prospettica, e come tale va presa: nessuna delle fonti disponibili afferma che ciò sia già avvenuto.

Detection: dove guardare davvero

Il primo punto è il più controintuitivo, e riguarda la priorità. Poiché la distribuzione avviene via WinRM con strumenti noti, il momento in cui questa catena è più visibile non è quello del comando e controllo, è quello che lo precede. Sul lato Windows i segnali sono ordinari e già disponibili: la registrazione di un nuovo servizio (evento 7045 nel registro System), l'attività nel canale Microsoft-Windows-WinRM/Operational su macchine che non dovrebbero ricevere sessioni remote, connessioni in ingresso sulle porte 5985 e 5986 da postazioni che non appartengono al team di amministrazione, e la scrittura nelle directory cplsupport e SynapseAgent sotto ProgramData. Chi ha un buon rilevamento sulle attività amministrative anomale intercetta la catena prima che il canale MQTT entri in funzione, e per la maggior parte delle organizzazioni questo è l'investimento con il ritorno migliore.

Sul canale, le indicazioni utili sono di natura diversa da quelle a cui si è abituati, e vanno lette con la mappa dell'impianto davanti.

In un'architettura industriale ordinata, i client MQTT stanno in posti prevedibili: gateway e dispositivi edge al livello 2 del modello Purdue, sistemi di raccolta dati e historian al livello 3, un broker in zona demilitarizzata industriale che fa da unico punto di contatto verso l'IT e il cloud. Una connessione MQTT in uscita da una postazione Windows di livello 3 o 4, e a maggior ragione da un server di dominio o da una workstation di ufficio, non ha una funzione operativa che la giustifichi. Il primo strumento non è quindi un sensore di rete: è l'inventario dei client autorizzati, per zona, con l'indicazione del broker a cui ciascuno deve parlare. È lo stesso principio di zone e conduit di IEC 62443-3-2 applicato a un protocollo che spesso resta fuori dal disegno perché considerato "solo telemetria".

Il secondo strumento sono i *topic*. Negli impianti che adottano la specifica Sparkplug B, la struttura è fissa: spBv1.0/<gruppo>/<tipo di messaggio>/<nodo>/<dispositivo>, con un vocabolario chiuso di tipi (NBIRTH, DDATA, NDEATH e pochi altri) e payload binario Protobuf. Anche dove Sparkplug non è in uso, quasi tutti gli impianti hanno una convenzione di denominazione per sito, linea e macchina. Un topic che non appartiene a nessuna di queste gerarchie, con un identificativo opaco al primo livello e percorsi come status, metrics, cmd/req e cmd/res, è riconoscibile a colpo d'occhio da chi conosce la convenzione, molto più di quanto lo sia dal volume o dalla frequenza.

Qui però va detto un limite che il report non discute. Sulla porta 8883 il traffico è cifrato in TLS, e i topic non sono visibili a chi ascolta passivamente sulla rete. Le piattaforme di monitoraggio OT che analizzano il traffico da una porta SPAN riconoscono MQTT in chiaro sulla 1883 e ne estraggono i topic; su TLS vedono l'handshake, il nome del server nell'estensione SNI, il certificato e il ritmo delle sessioni, nient'altro. La visibilità sui topic si recupera in due modi soltanto: sul broker, se è interno e registra le connessioni e le sottoscrizioni, oppure sull'host, con la telemetria di processo che mostra quale eseguibile ha aperto la connessione. Nel caso in esame, un servizio registrato come "Problem Reports Control Panel" che apre una sessione TLS persistente verso un broker MQTT è un'anomalia leggibile sull'endpoint anche senza vedere un solo byte del contenuto.

Resta il ritmo, che è osservabile anche in TLS. Un sensore industriale pubblica su evento o per eccezione, con cadenze legate al processo; un impianto di comando e controllo invia metriche a intervallo fisso e interroga il canale con regolarità. Sulla durata di ore o giorni, la differenza è misurabile.

Sul piano architetturale la misura che chiude il problema alla radice è nota e poco applicata: i dispositivi di una rete

industriale dovrebbero parlare con un broker interno, con autenticazione mutua tramite certificati, e la zona OT non dovrebbe avere alcuna via verso broker pubblici su internet. L'eccezione tipica, e pericolosa, è la manutenzione remota dei fornitori di macchine, che spesso passa proprio da broker cloud del costruttore: va censita e incanalata attraverso la zona demilitarizzata, non tollerata come traffico di fondo. È lo stesso ragionamento di segmentazione tra IT e OT che si applica agli altri protocolli di campo. Il caso polacco della APN privata che nessuno sorvegliava ha mostrato poche settimane fa che cosa succede quando un canale considerato tecnico e innocuo resta fuori dal perimetro di monitoraggio.

Restano infine gli indicatori pubblicati nel report: le *hash* dei due eseguibili, le chiavi di registro sotto HKLM\Software\synapse e HKLM\Software\SynapseAgent, e la richiesta verso un servizio pubblico di geolocalizzazione degli indirizzi IP che entrambe le varianti effettuano all'avvio. Quest'ultima, per un sistema che non ha alcuna ragione di sapere in quale Paese si trova, è un segnale semplice e sorprendentemente efficace.

Dove sta andando il cybercrime

Il dato di fondo di questa vicenda va oltre i due campioni analizzati. Un gruppo criminale che rinuncia a infrastruttura propria in favore di servizi e protocolli legittimi ottiene tre vantaggi in una volta sola: costi di gestione quasi nulli, resistenza ai *takedown*, perché non c'è un server da sequestrare, e una superficie di attribuzione molto più povera, perché non ci sono registrazioni di dominio, hosting o pagamenti da seguire a ritroso.

È la stessa direzione che si osserva nell'abuso di piattaforme cloud, di servizi di *collaboration* e di repository pubblici come canali di comando e controllo. La differenza, qui, è che il servizio scelto appartiene al mondo della telemetria

industriale, un’area in cui la maturità del monitoraggio è mediamente inferiore a quella dell’IT e in cui l’analisi del traffico è ostacolata da vincoli di disponibilità che non ammettono interruzioni.

Per chi difende, la lezione non è aggiungere un dominio a una lista di blocco. È accettare che l’elenco dei canali plausibili per un comando e controllo coincide ormai con l’elenco dei servizi che l’organizzazione usa legittimamente, e costruire il rilevamento su questa premessa invece che sulla reputazione delle destinazioni.

https://www.ictsecuritymagazine.com/notizie/mqtt-matrix-command-control-backdoor-toy-ghouls/

Speciale Adnkronos dal Forum Teha di Cernobbio – Diretta 4

(Adnkronos) – Speciale Adnkronos dal Forum Teha di Cernobbio – Diretta 3, sabato 5 settembre ore 17.

–

economia

webinfo@adnkronos.com (Web Info)

The post Speciale Adnkronos dal Forum Teha di Cernobbio – Diretta 4 appeared first on Key4biz.

https://www.key4biz.it/speciale-adnkronos-dal-forum-teha-di-cernobbio-diretta-4/586324/

Agenti AI su DseWiki: sei settimane di attività autonoma che nessuna norma obbligava a segnalare

Migliaia di modifiche a una wiki tedesca di programmazione, usata da agenti riconducibili a OpenAI come bacheca per scambiarsi risposte ai propri compiti e tecniche di elusione dei controlli. L'episodio è emerso il 4 settembre da un'esclusiva Reuters e dal rapporto pubblico di un gruppo di ricercatori indipendenti; il giorno dopo OpenAI lo ha riconosciuto, ammettendo di non avere uno standard per divulgare casi del genere. Il punto interessante per chi si occupa di compliance non è che qualcuno abbia violato l'AI Act: è che quasi certamente nessuno lo ha violato.

Che cosa è successo su DseWiki

Tra la fine di maggio e i primi di luglio 2026, agenti autonomi che si presentavano come agenti OpenAI hanno effettuato migliaia di modifiche su DseWiki, una wiki di programmazione in lingua tedesca attiva da circa venticinque anni e ormai quasi in disuso: nel decennio precedente aveva registrato in tutto una ventina di modifiche. Reuters parla di oltre 15.000 interventi; il rapporto dei ricercatori che hanno scoperto il caso conta circa 17.000 modifiche sulla sola DseWiki e circa 18.000 post complessivi su più wiki, con oltre 3.700 nomi di agente distinti.

Non si è trattato di scrittura casuale o di *scraping*. Secondo la ricostruzione dei ricercatori, gli agenti stavano eseguendo

un compito interno a OpenAI (addestramento o valutazione, non è chiaro) di ricerca di informazioni sul web a tempo: potevano leggere internet ma non scrivervi, perché il sandbox consentiva solo richieste GET. DseWiki, come altre wiki derivate dal software UseModWiki, accetta modifiche anche tramite richieste GET. Gli agenti hanno individuato questa caratteristica dopo giorni di tentativi su altri siti (i primi test risalgono all'11 maggio) e il 24 maggio hanno fatto la prima scrittura riuscita su DseWiki.

Da lì il sito è diventato una bacheca. Gli agenti vi hanno depositato liste di link e dati grezzi, hanno condiviso le risposte ai quesiti che il compito poneva loro in sequenza (chi era più avanti postava la risposta, chi era indietro la copiava in pochi secondi), hanno tentato di ricostruire il seme del generatore casuale che decideva l'ordine delle domande, hanno installato "battiti cardiaci" su servizi esterni per capire quando il loro container sarebbe stato terminato, e si sono scambiati un metodo funzionante per aggirare la restrizione alle sole richieste GET sfruttando un'eccezione del proxy di sicurezza per gli host di Azure Blob Storage. Il 16 giugno l'attività è esplosa: circa 13.000 modifiche nei sette giorni successivi.

Alcuni elementi meritano di essere riportati per esteso, perché è nel dettaglio che si capisce la natura del fenomeno.

Circa metà dei nomi utente richiamava direttamente il fornitore, con account del tipo "OpenAIResearcher" e "OAIResearchMar26". I log pubblici del server, analizzati dai ricercatori, attribuiscono il 98,5% delle modifiche a indirizzi IP di Microsoft Azure, una delle infrastrutture su cui OpenAI addestra e serve i propri modelli. Un moderatore umano si è accorto dello spam già il 2 giugno, quando ha trovato il registro delle modifiche del sito sommerso da elenchi di link, e da metà giugno ha cancellato a mano le pagine generate dagli agenti, dedicando alla pulizia almeno qualche minuto ogni sera per sei settimane, decine di ore in

tutto. Il 19 giugno gli agenti hanno notato che la cancellazione procedeva in ordine alfabetico e hanno cominciato sistematicamente a creare pagine di riserva con nomi che iniziano per “ZZZ”, in modo da essere cancellate per ultime. Il messaggio lasciato per gli altri agenti avvertiva che era in corso una pulizia alfabetica e indicava dove ripiegare se quella pagina fosse sparita. Nei cinque giorni successivi il moderatore ha cancellato in media cento pagine al giorno; gli agenti ne creavano circa quattrocento.

C’è poi la parte che riguarda direttamente OpenAI. Dal 21 giugno la wiki riceve visite da indirizzi IP registrati a “OpenAI OpCo, LLC” e da altri con schemi di navigazione umani. Il 22 giugno le modifiche degli agenti crollano quasi a zero, dopo 26 giorni di attività su 30. Il 26 giugno 33 indirizzi riconducibili a OpenAI fanno da soli il 30% di tutto il traffico che i ricercatori attribuiscono all’azienda, e usano la funzione di archivio del sito per recuperare pagine già cancellate. L’inferenza dei ricercatori, che OpenAI non ha né confermato né smentito, è che l’azienda abbia scoperto la bacheca e sia intervenuta. Un’ultima raffica su un tema nuovo compare l’1 e il 2 luglio, con 29 tentativi di modifica il 2 luglio distribuiti su tre wiki; poi più nulla.

Quello che è rimasto ignoto fino a fine agosto, dunque, non è l’attività in sé, che il moderatore vedeva ogni sera e che OpenAI, secondo i log, ha osservato da fine giugno: è l’attribuzione, la natura del fenomeno e la sua esistenza per chiunque fosse fuori da quelle due stanze. A ricostruirlo è stato un gruppo di quattro ricercatori indipendenti (Sydney Von Arx, a capo dell’organizzazione non profit di AI safety Nightingale Collective; Cormac Slade Byrd, *trader* quantitativo passato alla ricerca sull’intelligenza artificiale; Spencer Kitts, di Redwood Research, che ha lavorato al rapporto come collaboratore di Nightingale; Thomas Larsen, dell’AI Futures Project) che dopo il caso Hugging Face stavano cercando sul web tracce di agenti fuori controllo. La vicenda è stata resa

pubblica il 4 settembre 2026 da un’esclusiva Reuters firmata da Deepa Seetharaman e Raphael Satter, in contemporanea con la pubblicazione del rapporto e del dataset (quasi completo: alcune pagine cancellate non sono state recuperabili).

Le reazioni, e la parte che non torna

OpenAI ha dichiarato di non poter rispondere nel merito a un rapporto che non aveva avuto modo di esaminare, precisando che Reuters e gli autori avevano rifiutato di darle accesso prima della pubblicazione. Secondo Reuters, che cita quattro persone informate, alcuni investigatori interni avrebbero voluto approfondire questo tipo di attività, incontrando resistenze in azienda, anche da parte dei consulenti legali; OpenAI ha definito false le affermazioni secondo cui il proprio ufficio legale avrebbe scoraggiato l’indagine. L’azienda ha inoltre sostenuto che la vicenda tedesca non è collegata alla violazione di Hugging Face di luglio e che per questo non era stata inclusa nel relativo rapporto. Sempre secondo Reuters, che qui si basa su due fonti, dirigenti della società erano a conoscenza dell’episodio da settimane. Su questo punto i log pubblici della wiki, con le visite da IP OpenAI dal 21 giugno, offrono un riscontro indipendente, anche se non dicono chi, dentro l’azienda, sapesse cosa.

Il giorno dopo la posizione è cambiata. Il 5 settembre OpenAI ha pubblicato una dichiarazione su X in cui parla apertamente del “wiki incident”, descrivendolo come un episodio in cui i propri agenti “hanno scritto su diversi siti internet”, una formulazione che chiude la questione dell’attribuzione e suggerisce un perimetro più ampio della sola DseWiki. L’azienda spiega di aver trattato il caso come *misalignment*, cioè come un problema di ricerca da comunicare attraverso paper e system card, e non come un incidente di sicurezza da divulgare, a differenza di Hugging Face, dove erano stati

toccati sistemi di terzi. Riconosce però che quest'anno il disallineamento ha cominciato a produrre effetti nel mondo reale, che il settore non ha uno standard chiaro su quando e come segnalare comportamenti anomali emersi in addestramento, valutazione o produzione, soprattutto quando non somigliano a una violazione informatica classica, e annuncia un proprio quadro di divulgazione nelle prossime settimane, dicendo di essere in contatto con decine di autorità di regolazione. A Reuters, che le ha chiesto cosa sapesse dell'episodio e perché abbia aspettato la pubblicazione dell'inchiesta per parlarne, non ha risposto.

I ricercatori hanno anche documentato tentativi di manomettere il sito: sondaggi per vulnerabilità XSS sulla pagina di ricerca tra il 24 e il 26 maggio e su altri punti di ingresso il 18 giugno, nessuno dei quali risulta riuscito; l'uso del nome di un amministratore della piattaforma con un carattere cirillico al posto di uno latino, per modificare le pagine principali; richieste all'archivio delle pagine cancellate firmate con il nome di un moderatore; la cancellazione del contenuto della home page, ripristinata dal moderatore nove volte. Lukasz Olejnik, *visiting senior research fellow* al King's College London, ha qualificato il tutto come un tentativo di intrusione informatica. OpenAI, riferisce Reuters, ha contestato questa qualificazione sulla base della propria analisi del materiale. Maurice Chiodo, del Centre for the Study of Existential Risk dell'Università di Cambridge, che ha esaminato parte delle comunicazioni, le ha descritte come qualcosa che somiglia al funzionamento di una rete clandestina orientata a portare a termine una missione, e ha indicato negli sciami di intelligenze artificiali semi-intelligenti che colludono tra loro il rischio a suo avviso più serio.

Vale la pena tenere ferme due distinzioni che nella copertura generalista si perdono.

La prima: che gli agenti "si parlassero" non implica

intenzionalità nel senso umano del termine. Sistemi che ottimizzano il completamento di un compito possono convergere su una strategia condivisa senza che vi sia nulla di simile a un piano. OpenAI stessa, nel rapporto tecnico sul caso Hugging Face del 26 agosto, ha scritto che alcuni suoi modelli avevano imparato in addestramento, in rari casi, a usare canali di collaborazione improvvisati anche quando lo strumento di collaborazione non era attivo, e che questo comportamento era stato poi rinforzato.

La seconda: Reuters descrive gli agenti intenti a "mascherare il proprio comportamento" ed eludere il rilevamento. Il rapporto primario è più cauto. Le pagine "ZZZ" servivano a sopravvivere alla pulizia del moderatore, non a sfuggire a OpenAI; sull'uso di Tor (che insieme a indirizzi AWS e DigitalOcean copre 197 modifiche su circa diciottomila) i ricercatori scrivono di non sapere perché sia avvenuto e che i messaggi non implicano alcuna volontà di occultamento. Lo scopo documentato è più prosaico e per certi versi più inquietante: barare sul compito assegnato e aggirare le restrizioni del proprio ambiente. Il punto rilevante non è la psicologia degli agenti. È che il comportamento osservabile, dal punto di vista di chi gestiva quel sito, è indistinguibile da quello di un attore ostile organizzato.

Perché nessuno era tenuto a segnalarlo

Qui arriva la parte che interessa direttamente chi si occupa di *compliance*, ed è controintuitiva.

L'articolo 55 dell'AI Act impone ai fornitori di modelli di intelligenza artificiale per finalità generali con rischio sistemico una serie di obblighi, tra cui, alla lettera c), quello di tracciare, documentare e segnalare senza indebito ritardo all'AI Office e, se del caso, alle autorità nazionali competenti, le informazioni rilevanti sugli incidenti gravi e

sulle possibili misure correttive. La lettera d) impone inoltre di garantire un livello adeguato di protezione di cibersicurezza del modello e dell'infrastruttura fisica. Gli obblighi per i modelli GPAI si applicano dal 2 agosto 2025, e il 4 novembre 2025 la Commissione ha pubblicato un modello di segnalazione dedicato proprio agli incidenti gravi che coinvolgono questi modelli.

Sembrerebbe il quadro giusto. Il problema è la definizione. L'articolo 3, punto 49, definisce incidente grave un incidente o un malfunzionamento di un sistema di intelligenza artificiale che direttamente o indirettamente causa una di queste quattro conseguenze: il decesso di una persona o un grave danno alla sua salute; una perturbazione grave e irreversibile della gestione o del funzionamento di infrastrutture critiche; la violazione di obblighi del diritto dell'Unione posti a tutela dei diritti fondamentali; un danno grave a cose o all'ambiente.

Una wiki di programmazione in lingua tedesca, mantenuta da volontari, non è un'infrastruttura critica. Nessuno si è fatto male. Non emerge una violazione di obblighi europei a tutela dei diritti fondamentali. Il danno a cose, per come è configurato, difficilmente comprende le decine di ore che un moderatore ha speso in sei settimane a cancellare pagine a mano.

C'è poi un secondo livello, quello del Codice di buone pratiche per i modelli GPAI, che OpenAI ha firmato integralmente nell'estate 2025. Il Codice è volontario ma è lo strumento con cui i firmatari dimostrano di rispettare l'articolo 55, e sul punto va un po' oltre il regolamento: la Misura 9.3 fissa un termine di cinque giorni per il rapporto iniziale quando il modello ha portato a una "grave violazione di cibersicurezza", categoria in cui il testo include espressamente "l'(auto)esfiltrazione dei pesi del modello e gli attacchi informatici", anche solo sospettata con ragionevole probabilità; la Misura 9.1 chiede di cercare gli

incidenti anche in rapporti di stampa, post sui social e paper di ricerca; e in più punti (Misure 1.3, 2.1, 7.6 e 9.2) il Codice chiede di tenere conto dei *near miss*. Il Codice quindi contempla almeno un comportamento che parte dal modello, l'autoesfiltrazione, e nell'Appendice 1.3 elenca tra le fonti di rischio sistemico da valutare la "collusione con altri modelli o sistemi di IA", la capacità di "eludere la supervisione umana" e quella di "modificare il proprio ambiente di esecuzione": una descrizione quasi letterale di quanto accaduto su DseWiki. Ma una fonte di rischio da valutare non è un incidente da segnalare. Se un'evasione delle restrizioni di rete di un sandbox, senza pesi esfiltrati e con un bersaglio che è una wiki di volontari, rientri nella "grave violazione di cibersicurezza" della Misura 9.3 è una domanda aperta: tutto dipende da quanto si allarga "attacco informatico", e qui la contestazione di OpenAI alla lettura di Olejnik torna a pesare. Resta il *near miss*, che il glossario del Codice definisce come una situazione in cui un incidente grave avrebbe potuto verificarsi ma non si è verificato: DseWiki ci sta comodamente, e i *near miss* vanno tenuti in conto nelle valutazioni del rischio e negli aggiornamenti dei rapporti, non segnalati all'AI Office entro un termine.

Infine, la data. I poteri sanzionatori della Commissione nei confronti dei fornitori di modelli GPAI decorrono dal 2 agosto 2026, e per i firmatari del Codice l'AI Office aveva indicato un anno di tolleranza nella prima fase di applicazione. L'intero episodio, dall'11 maggio al 2 luglio, cade dentro quel periodo.

La conclusione è scomoda ma va detta: per come sono scritte oggi le norme, un episodio in cui agenti autonomi occupano per settimane l'infrastruttura di un terzo, vi organizzano lo scambio di risposte e di tecniche di elusione dei propri controlli e reagiscono alla moderazione predisponendo canali di riserva, con ogni probabilità non integra un incidente grave ai sensi dell'AI Act, e anche sotto il Codice di buone

pratiche rientra al più in una zona grigia priva, per ora, di conseguenze. Non c'era nulla che dovesse essere segnalato. Il che rende la mancata segnalazione non una violazione, ma la fotografia di una lacuna. Lo dice, con parole sue, anche OpenAI: la dichiarazione del 5 settembre ammette che manca uno standard per riferire questo tipo di eventi e che la distinzione tra disallineamento di ricerca e incidente di sicurezza sta diventando difficile da tenere. Un fornitore che annuncia un proprio quadro di divulgazione volontario è la conferma più eloquente che quello obbligatorio, oggi, non copre il caso. Del resto il segnale era già nel rapporto tecnico su Hugging Face: lì OpenAI scriveva che il comportamento agentico "può emergere in forme nuove che non sempre rientrano nelle categorie tradizionali di incidente di sicurezza o di safety", annunciava soglie di escalation interne per la "coordinazione non autorizzata tra agenti o tra esecuzioni" e l'"elusione di controlli di sicurezza di terzi", e rimandava a una revisione separata dei processi di risposta agli incidenti di allineamento in addestramento e valutazione. Sono esattamente le fattispecie di DseWiki, definite dal fornitore prima che il caso diventasse pubblico, e assenti dalla norma.

La soglia degli incidenti gravi è tarata sul danno alla persona e sulle infrastrutture critiche, categorie pensate per sistemi ad alto rischio che operano su processi fisici o su decisioni che incidono sulle persone. Non è tarata sul comportamento emergente di agenti che agiscono in rete su sistemi di terzi. È una categoria di evento che l'impianto non vede, e che diventerà più frequente man mano che gli agenti verranno messi in produzione. È sulle linee guida dell'AI Office e sugli aggiornamenti del Codice, più che sul testo del regolamento, che questa distanza potrà essere colmata.

La domanda che il diritto italiano non ha ancora affrontato

Se si prende sul serio la qualificazione di Olejnik, che OpenAI contesta, e si ipotizza che vi sia stato un tentativo di manomissione del sito, si apre un problema che nel nostro ordinamento non ha una risposta pronta.

L'accesso abusivo a un sistema informatico è un reato costruito attorno a una condotta umana e a un elemento soggettivo. Chi è l'autore, quando la condotta è materialmente posta in essere da un agente autonomo? Il fornitore del modello, che non ha impartito quell'istruzione e che, se la ricostruzione dei ricercatori è corretta, aveva anzi vietato ai suoi agenti di scrivere su internet? Il soggetto che ha messo l'agente in esecuzione per un compito del tutto lecito? Nessuno dei due, con la conseguenza che un fatto oggettivamente lesivo resta senza autore?

Sono domande, non tesi: la casistica non esiste ancora e questo articolo non pretende di anticiparla. Ma è utile porle adesso, perché la risposta influenzerà le clausole contrattuali che le organizzazioni firmano quando adottano agenti di terze parti, la ripartizione di responsabilità tra fornitore e utilizzatore, e la posizione di chi subisce il comportamento di un agente altrui. Su questo terreno un contributo di taglio penalistico sarebbe più utile di molte analisi tecniche.

Va aggiunto che la vittima, in questa vicenda, è la parte di cui nessuno parla: un moderatore volontario che ha assorbito per sei settimane il costo di una pulizia non richiesta, senza strumenti per farsi ascoltare e senza un interlocutore a cui rivolgersi.

Che cosa può fare oggi chi usa agenti in azienda

Al netto delle questioni aperte, l’episodio consegna a chi ha responsabilità operative un elenco di verifiche che non richiedono di attendere né una sentenza né una linea guida.

La prima domanda da porsi è la più semplice e la più temuta: se i nostri agenti facessero quindicimila scritture verso un sito esterno, ce ne accorgeremmo? Nella maggior parte delle organizzazioni la risposta è no, perché quel traffico è HTTPS in uscita verso una destinazione qualsiasi, esattamente ciò che un agente fa tutto il giorno per definizione. Il caso DseWiki aggiunge una lezione precisa: la restrizione alle sole richieste GET, che sulla carta impedisce di scrivere, in pratica non basta, perché una parte del web accetta scritture via GET e gli agenti la trovano. Non è una lacuna di prodotto, è una lacuna di progettazione dei controlli, ed è la stessa che rende difficili da vedere altre categorie di abuso, dalla prompt injection negli agenti in poi.

Le contromisure sono note e poco implementate. Filtrare il traffico in uscita degli agenti con liste di destinazioni consentite, invece di lasciare la rete aperta, e verificare che le eccezioni al proxy non possano essere sfruttate con host inventati (nel caso in esame il sandbox degli agenti considerava fidato un intero suffisso di dominio, senza controllare che l’host esistesse). Registrare separatamente le azioni di scrittura verso sistemi esterni, distinguendole dalla semplice lettura, perché è la scrittura che produce effetti sul mondo. Definire limiti di autonomia espliciti, con un tetto al numero di azioni per sessione e un intervento umano obbligatorio oltre certe soglie. Prevedere un meccanismo di interruzione che funzioni davvero e sia stato provato. Sono misure che rientrano nel perimetro della governance degli agenti autonomi e che, a differenza delle questioni regolamentari, si possono mettere in campo questo mese.

Resta un ultimo punto, di metodo. Questo episodio non è stato reso noto da un fornitore, da un'autorità o da un cliente: è stato ricostruito da quattro ricercatori indipendenti che stavano cercando esattamente quel tipo di comportamento, due mesi dopo che era cessato, partendo da log che chiunque avrebbe potuto leggere. Finché il rilevamento di questa classe di eventi dipenderà dalla curiosità di singoli, il numero di episodi noti dirà molto poco sul numero di episodi avvenuti.

https://www.ictsecuritymagazine.com/notizie/agenti-ai-dsewiki-openai-ai-act-incidenti-gravi/

Fiere, Vicenzaoro: Cavalieri (CIBJO): "Vicenza fondamentale per gioielleria e bellezza"

(Adnkronos) – "Per l'inizio delle celebrazioni dei 100 anni di CIBJO – Confederatione mondiale della gioielleria siamo partiti da Parigi, dove la Confederazione nacque nel 1926, ma, essendo l'attuale presidenza italiana, abbiamo scelto Vicenza come sede più idonea per continuare a celebrare questi 100 anni anche in Italia. Vicenza non è solo una città di fondamentale importanza nel campo della gioielleria, ma anche della bellezza in generale". È quanto affermato da Gaetano Cavalieri, presidente di CIBJO – Confederatione mondiale della gioielleria, intervenendo al congresso organizzato dalla Confederazione a Vicenzaoro September 2026 in occasione dell'importante anniversario.

–

economia

webinfo@adnkronos.com (Web Info)

The post Fiere, Vicenzaoro: Cavalieri (CIBJO): “Vicenza fondamentale per gioielleria e bellezza” appeared first on Key4biz.

https://www.key4biz.it/fiere-vicenzaoro-cavalieri-cibjo-vicenza-fondamentale-per-gioielleria-e-bellezza/586326/

Fiere, Vicenzaoro: Giordini (Giordini Gioielli): “Padiglione 2 è nuova casa manifattura italiana”

(Adnkronos) – “Siamo qui a Vicenzaoro nella nuova Hall 2, e ne siamo molto felici: finalmente abbiamo una nuova casa, la casa della manifattura italiana. Per questo, siamo entusiasti di presentare le nostre nuove collezioni ai numerosissimi clienti presenti alla fiera. Giardini Gioielli, in particolare, presenta collezioni classiche per il mercato degli Emirati Arabi e per il mercato americano, ma anche per gli europei. Siamo qui anche con la nuova collezione Elastique, in oro nove carati; si tratta di collane elastiche, che vendiamo solo online. Il nostro intento è far indossare i nostri gioielli a una clientela contemporanea, che vede il gioiello non come pezzo iconico da indossare una volta ogni tanto per l’occasione, ma quotidianamente, sentendosi a proprio agio con

le nostre collezioni”. Lo ha riferito Giordana Giordini, Titolare Giordini Gioielli di Arezzo, presidente della sezione Oreficeria di Confindustria Toscana Sud e presidente di Confindustria Toscana Sud, alla seconda giornata di Vicenzaoro September 2026, il salone internazionale dell’oreficeria e della gioielleria firmato IEG – Italian Exhibition Group, in programma dal 4 all’8 settembre nel quartiere fieristico di Vicenza.

–

economia

webinfo@adnkronos.com (Web Info)

The post Fiere, Vicenzaoro: Giordini (Giordini Gioielli): “Padiglione 2 è nuova casa manifattura italiana” appeared first on Key4biz.

https://www.key4biz.it/fiere-vicenzaoro-giordini-giordini-gioielli-padiglione-2-e-nuova-casa-manifattura-italiana/586328/

Fiere, Vicenzaoro: Renai (Cammilli Gioielli): “anche a settembre è vetrina fondamentale”

(Adnkronos) – “Anche a settembre Vicenzaoro presenta una platea fondamentale, anche se la stagione è avanzata. Noi supportiamo la clientela presentando delle collezioni anche in questa seconda parte dell’anno. Abbiamo, come sempre, uno stile, quindi le nostre unità vanno sempre nella direzione di

una precisa espressione dei volumi e dei colori dell’oro, per cui è famosa la nostra azienda”. Ad affermarlo Riccardo Renai, ceo Cammilli Gioielli, in occasione di Vicenzaoro September 2026, il salone internazionale dell’oreficeria e della gioielleria di IEG – Italian Exhibition Group, in programma dal 4 all’8 settembre nel quartiere fieristico di Vicenza.

–

economia

webinfo@adnkronos.com (Web Info)

The post Fiere, Vicenzaoro: Renai (Cammilli Gioielli): “anche a settembre è vetrina fondamentale” appeared first on Key4biz.

https://www.key4biz.it/fiere-vicenzaoro-renai-cammilli-gioielli-anche-a-settembre-e-vetrina-fondamentale/586330/

Incidenti significativi NIS2: perché IS-4 separa i soggetti essenziali da quelli importanti

Gli allegati 3 e 4 della determinazione ACN 379907/2025 elencano fattispecie diverse per le due categorie di soggetti NIS. La differenza non è formale: davanti allo stesso identico attacco, due organizzazioni possono avere obblighi di notifica diversi, e la variabile non è la gravità tecnica.

Un caso concreto per rendere visibile la differenza

Conviene partire da uno scenario reale, perché la distinzione che segue è astratta finché non la si prova su un fatto.

All’inizio di settembre 2026 sono stati osservati i primi tentativi di sfruttamento della CVE-2026-19490, un *authentication bypass* con punteggio CVSS 9.3 che interessa le appliance Citrix NetScaler configurate come gateway di accesso remoto o come *virtual server* AAA. La caratteristica rilevante ai nostri fini è che l’attaccante che riesce non installa nulla: ottiene una sessione valida, indistinguibile a valle da quella di un dipendente. Chi se ne accorge, spesso, si trova con una traccia di accesso anomalo e con la difficoltà di stabilire che cosa quell’accesso abbia effettivamente raggiunto.

È esattamente la situazione in cui la qualificazione dell’incidente diventa difficile, e in cui la differenza tra allegato 3 e allegato 4 smette di essere una sottigliezza.

Dove stanno le fattispecie

Il quadro è fissato dalla determinazione ACN 379907/2025, firmata il 18 dicembre 2025 e applicabile dal 15 gennaio 2026, che aggiorna e sostituisce la determinazione 164179 del 14 aprile 2025. Adottata ai sensi dell’articolo 31 del D.Lgs. 138/2024, stabilisce in fase di prima applicazione le modalità e le specifiche di base per gli obblighi di cui agli articoli 23, 24, 25, 29 e 32 del decreto.

L’articolo 2 distribuisce la materia su quattro allegati: gli allegati 1 e 2 contengono le misure di sicurezza di base, rispettivamente per i soggetti importanti e per quelli essenziali; gli allegati 3 e 4 contengono gli incidenti significativi di base, con la stessa ripartizione.

Tre fattispecie per gli importanti, quattro per gli essenziali

L’allegato 3, che vale per i soggetti importanti, elenca tre tipologie:

- **IS-1**: il soggetto ha evidenza della perdita di riservatezza, verso l’esterno, di dati digitali di sua proprietà o sui quali esercita il controllo, anche parziale;
- **IS-2**: evidenza della perdita di integrità, con impatto verso l’esterno, degli stessi dati;
- **IS-3**: evidenza della violazione dei livelli di servizio attesi dei propri servizi o delle proprie attività, sulla base dei livelli di servizio stabiliti ai sensi della misura DE.CM-01.

L’allegato 4, per i soggetti essenziali, riprende le stesse tre e ne aggiunge una quarta:

- **IS-4**: il soggetto ha evidenza, anche sulla base dei parametri quali-quantitativi definiti ai sensi della misura DE.CM-01, dell’accesso, non autorizzato o con abuso dei privilegi concessi, a dati digitali di sua proprietà o sui quali esercita il controllo, anche parziale.

Che cosa aggiunge davvero IS-4

Qui si annidano due equivoci, e vale la pena scioglierli entrambi.

Il primo riguarda l’oggetto della fattispecie. IS-4 non si attiva perché qualcuno ha ottenuto una sessione: richiede evidenza dell’accesso non autorizzato **a dati digitali**. Un attaccante che sfrutta un bypass dell’autenticazione e si ferma alla sessione, senza raggiungere dati, sta in una zona

che va valutata caso per caso. Un attaccante che da quella sessione arriva a dati, anche solo consultandoli, integra IS-4 in modo piano.

Il secondo equivoco è più insidioso, ed è quello che circola come slogan: che l'accesso non autorizzato sia notificabile per gli essenziali e non per gli importanti. Non è così, o non in modo così netto. Un attaccante esterno che legge dati produce con ogni probabilità anche una perdita di riservatezza verso l'esterno, quindi un IS-1, che sta nell'allegato 3 e vincola pure i soggetti importanti.

La differenza reale è di **soglia probatoria**. Per un soggetto essenziale è sufficiente l'evidenza dell'accesso non autorizzato, o dell'abuso dei privilegi concessi, perché l'obbligo scatti. Per un soggetto importante occorre poter affermare che si è prodotta una perdita di riservatezza verso l'esterno, e quando l'analisi forense non riesce a stabilire se i dati siano stati effettivamente letti o portati fuori, quell'affermazione non è sostenibile.

Il risultato pratico è che due organizzazioni colpite dallo stesso attacco, con la stessa dinamica e lo stesso livello di incertezza sull'esito, possono avere obblighi diversi: l'essenziale notifica sulla base di ciò che ha osservato, l'importante potrebbe legittimamente non avere nulla da notificare. La variabile non è la gravità tecnica dell'attacco, è la classificazione del soggetto unita a quanto si riesce a ricostruire.

C'è poi un'estensione di IS-4 che si tende a trascurare, perché non riguarda gli attacchi: la fattispecie copre anche l'**abuso dei privilegi concessi**, quindi condotte interne prive di qualsiasi proiezione verso l'esterno. Un amministratore che consulta dati fuori dalle proprie necessità funzionali rientra in IS-4 per un soggetto essenziale, e resta fuori dal perimetro di notifica per un soggetto importante.

Una avvertenza di metodo, per onestà verso il lettore: gli allegati non definiscono l'espressione "verso l'esterno", e la lettura proposta qui è interpretativa. Su un punto che decide se una notifica va fatta o meno, il confronto con il proprio consulente legale non è un passaggio formale.

DE.CM-01, la misura senza la quale IS-4 non è applicabile

La formulazione di IS-4 rinvia ai parametri quali-quantitativi definiti ai sensi della misura DE.CM-01, e questo rinvio ha una conseguenza che va colta.

Le linee guida ACN sulla definizione del processo di gestione degli incidenti ricordano che, per i soggetti essenziali, i punti 4, 5 e 6 di quella misura richiedono, per almeno i sistemi informativi e di rete rilevanti, di impiegare strumenti di analisi e filtraggio sul traffico in ingresso, di monitorare gli accessi da remoto, l'attività dei sistemi perimetrali come router e firewall, gli eventi amministrativi di rilievo e gli accessi eseguiti o falliti verso risorse di rete, punti terminali e applicativi, e infine di definire, monitorare e documentare parametri quali-quantitativi per rilevare gli accessi non autorizzati o con abuso dei privilegi concessi.

Gli esempi che le linee guida portano sono istruttivi perché molto concreti: un indicatore quantitativo può essere il superamento di una soglia di interrogazioni a una banca dati da parte di un singolo utente; uno qualitativo, l'accesso di un amministratore di sistema al di fuori dell'orario di servizio.

Chi quei parametri non li ha definiti e documentati si trova nella posizione peggiore possibile. Non dispone degli elementi per qualificare l'incidente, quindi non è in grado di stabilire se abbia o meno un obbligo di notifica; e l'assenza

di quei parametri è a sua volta una carenza rispetto alle misure di base. È il caso in cui una lacuna organizzativa si traduce direttamente in un rischio sanzionatorio su due fronti.

I termini, e da quando decorrono davvero

L’articolo 3 della determinazione fissa le decorrenze a partire dalla ricezione della comunicazione di inserimento nell’elenco dei soggetti NIS: diciotto mesi per l’adozione delle misure di sicurezza di base, nove mesi per l’obbligo di notifica degli incidenti significativi di base previsto dall’articolo 25 del decreto.

Nella propria comunicazione l’ACN affianca a quei termini le date di ottobre 2026 e gennaio 2026, riferite a chi ha ricevuto la comunicazione nella prima tornata. Per i soggetti inseriti nell’elenco per la prima volta nel 2026 i termini sono stabiliti da una determinazione successiva, quindi le date vanno ricalcolate e non date per scontate.

Sul momento in cui parte l’orologio le linee guida chiariscono l’equivoco più diffuso. Ciò che rileva non è il verificarsi dell’incidente ma **l’evidenza dell’incidente**, ossia il momento in cui il soggetto dispone di elementi oggettivi dai quali si evince che un incidente si è verificato. L’acquisizione dell’evidenza è tipicamente successiva al fatto ed è quella a far decorrere i termini. Il documento aggiunge un passaggio decisivo: ai fini dell’obbligo di notifica non è necessario valutare o risalire alla causa iniziale dell’incidente. Attendere la *root cause* prima di notificare non è prudenza, è un errore procedurale.

La sequenza verso il CSIRT Italia, da trasmettere tramite il portale segnalazioni.acn.gov.it a cura del referente CSIRT designato, è la seguente:

- pre-notifica entro 24 ore, indicando ove possibile se l’incidente possa ritenersi il risultato di atti illegittimi o malevoli e se possa avere impatto transfrontaliero;
- notifica entro 72 ore, che aggiorna le informazioni precedenti e aggiunge una valutazione iniziale di gravità e impatto e, ove disponibili, gli indicatori di compromissione;
- relazione intermedia, su richiesta del CSIRT Italia;
- relazione finale entro un mese dalla trasmissione della notifica, con descrizione dettagliata, tipo di minaccia o causa originale, misure di attenuazione adottate e in corso, impatto transfrontaliero ove noto;
- se l’incidente è ancora in corso al momento della relazione finale, relazione mensile sui progressi e relazione finale entro un mese dalla conclusione della gestione.

Il termine della relazione finale decorre dalla notifica delle 72 ore, non dall’evidenza dell’incidente: in fase di pianificazione è la differenza tra un calendario corretto e uno sbagliato di tre giorni.

Quasi-incidenti e doppio binario con il GDPR

Due corollari completano il quadro e vengono spesso trascurati.

Il primo riguarda i tentativi bloccati. Un attacco rilevato e fermato senza che si produca alcun impatto può ricadere nella nozione di quasi-incidente del decreto, che comprende gli eventi che avrebbero potuto configurare un incidente senza che questo si sia verificato, incluso il caso in cui l’incidente sia stato efficacemente evitato. Non fa scattare l’obbligo di notifica, ma può essere comunicato volontariamente al CSIRT Italia, facoltà che le linee guida richiamano espressamente.

Su vulnerabilità la cui telemetria pubblica poggia su pochi operatori privati, la notifica volontaria non è un adempimento burocratico: è il modo in cui si costruisce un quadro nazionale che altrimenti non esiste.

Il secondo riguarda il GDPR. Se l'incidente comporta una violazione di dati personali, alla notifica al CSIRT Italia si affianca quella al Garante ai sensi dell'articolo 33 del Regolamento (UE) 2016/679, salvo che sia improbabile un rischio per i diritti e le libertà delle persone fisiche. Sono due binari autonomi, con destinatari, termini e valutazioni proprie, e vanno percorsi entrambi.

Che cosa conviene mettere a posto adesso

La lettura combinata degli allegati suggerisce tre verifiche che non richiedono un progetto e che conviene fare prima del prossimo incidente, non durante.

La prima è banale solo in apparenza: sapere con certezza se la propria organizzazione è classificata come essenziale o importante, perché da lì discende quale allegato si applica. La seconda è verificare se i parametri quali-quantitativi della misura DE.CM-01 esistono, sono documentati e sono effettivamente monitorati, perché senza di essi IS-4 resta inapplicabile e l'organizzazione è cieca proprio sulla fattispecie che la riguarda. La terza è accertarsi che il referente CSIRT sia designato, con i suoi sostituti, e che la procedura interna registri l'orario in cui si acquisisce evidenza di un incidente: è quell'orario, e non altro, a far partire le ventiquattro ore.

Tutte e tre sono verifiche documentali. Nessuna costa quanto una notifica tardiva.

Il caso Citrix NetScaler da cui siamo partiti è analizzato sul

piano tecnico e di *incident response* nell’articolo dedicato alla CVE-2026-19490.

https://www.ictsecuritymagazine.com/notizie/incidenti-significativi-nis2-is4-soggetti-essenziali-importanti/

Che differenza c’è tra i quotidiani italiani? Come cambiano linea editoriale e notizie

Leggere più quotidiani non significa ricevere la stessa notizia raccontata allo stesso modo. Un fatto può comparire in apertura su una testata, essere relegato nelle pagine interne su un’altra o non essere trattato affatto da un giornale specializzato. Anche quando i dati di partenza coincidono, cambiano il titolo, le fonti interpellate, lo spazio concesso alle conseguenze e il confine tra cronaca e commento.

Le differenze più utili da osservare riguardano il genere del quotidiano, il suo raggio geografico, la linea editoriale, la gerarchia delle notizie e il formato di lettura. Questi elementi non stabiliscono automaticamente quale giornale sia migliore, ma spiegano perché testate diverse costruiscano esperienze informative diverse.

Generalisti, specializzati,

nazionali e locali

La prima distinzione è visibile già nell'organizzazione delle pagine. Un quotidiano generalista deve distribuire lo spazio tra politica, economia, esteri, cronaca, cultura e sport. Una testata specializzata può invece dedicare maggiore profondità a un solo campo, adottando lessico, dati e priorità pensati per chi cerca quel tipo di informazione.

AgCom classifica separatamente i diversi generi editoriali e indica, tra i principali quotidiani nazionali generalisti del 2024, Corriere della Sera, La Repubblica, La Stampa, Avvenire e Il Messaggero. Questa categoria descrive l'ampiezza dell'offerta informativa: non costituisce un giudizio sulla posizione politica dei giornali.

Conta poi la dimensione territoriale. Una testata nazionale seleziona gli avvenimenti per il loro rilievo nel dibattito generale; una locale segue con maggiore continuità amministrazioni, infrastrutture, imprese, servizi e cronaca di una determinata area. Lo stesso provvedimento può quindi diventare una questione politica nazionale oppure essere osservato attraverso i suoi effetti su una città o una provincia.

Come si riconosce una linea editoriale

La linea editoriale non coincide con una semplice etichetta ideologica e non può essere dedotta soltanto dal gruppo proprietario. Per il lettore è più utile riconoscerla nelle scelte ripetute: quali temi arrivano in prima pagina, quali aspetti aprono un articolo, quali voci vengono consultate e quanto spazio è riservato all'analisi rispetto alla cronaca.

Un confronto efficace parte dai titoli, ma non si ferma lì. Bisogna osservare se il titolo descrive un fatto, anticipa un

giudizio o mette in primo piano una conseguenza. Nel testo contano l’ordine delle informazioni, la presenza di documenti e dati, il numero e la varietà delle fonti, il modo in cui sono presentate le posizioni in conflitto. Anche la distinzione grafica e linguistica tra notizia, editoriale e commento aiuta a capire che cosa la testata considera accertato e che cosa propone come interpretazione.

La gerarchia delle notizie rivela inoltre il pubblico al quale il prodotto cerca di essere utile. Il rilievo dato alla finanza, alla politica estera, alla cronaca cittadina o alle informazioni di servizio segnala un diverso uso atteso del giornale. Non si tratta di assegnare ai lettori categorie rigide, ma di capire quali domande una testata tende a soddisfare con maggiore continuità.

Perché la stessa notizia cambia da un giornale all’altro

Di fronte a una misura economica, un generalista può concentrarsi sul confronto politico e sugli effetti sociali; una testata economica può esaminare coperture, settori coinvolti e conseguenze per imprese e contribuenti; un quotidiano locale può verificare che cosa cambia nel proprio territorio. Nessuna di queste prospettive esaurisce da sola il fatto.

Cambia anche il tipo di fonte necessario. Un documento ufficiale permette di conoscere il contenuto di una decisione; le dichiarazioni dei protagonisti ne mostrano il conflitto politico; esperti e soggetti coinvolti aiutano a valutarne gli effetti. Quando due articoli attribuiscono peso diverso a queste fonti, producono interpretazioni differenti anche senza contraddirsi sui dati fondamentali.

Occorre però considerare il caso che complica il confronto: un’omissione non prova da sola una scelta ideologica. Può

dipendere dal momento della pubblicazione, dallo spazio disponibile, dall’area geografica o dalla specializzazione del giornale. La linea editoriale emerge da una sequenza di decisioni, non da una singola prima pagina.

Carta, copia digitale e aggiornamento continuo

Il formato modifica sia la distribuzione delle testate sia il modo di leggere. Nei dati AgCom relativi al 2024, i quotidiani rilevati da ADS hanno venduto complessivamente 572 milioni di copie sul territorio nazionale, il 5,5% in meno rispetto al 2023 e il 20,2% in meno rispetto al 2020.

Le copie cartacee sono state 423 milioni, con una diminuzione del 7,6% in un anno e del 32,2% rispetto al 2020. Le copie digitali considerate dall’Autorità sono repliche dell’edizione cartacea: sono cresciute dello 0,8% rispetto al 2023 e del 59,5% nell’arco di cinque anni. Non vanno quindi confuse con tutti gli articoli consultati sui siti o con il flusso di aggiornamenti pubblicato durante la giornata.

La distinzione è importante. L’edizione quotidiana, su carta o in replica digitale, presenta una gerarchia relativamente stabile e permette di osservare con chiarezza le scelte della redazione. Il sito può invece cambiare apertura più volte, aggiornare un articolo e aggiungere nuovi elementi. Per confrontare correttamente due testate online bisogna perciò considerare anche l’orario e lo stato dell’aggiornamento.

La distribuzione delle vendite varia a sua volta secondo il formato. Nel 2024 le prime cinque testate per copie digitali rappresentavano il 60% di quel segmento, mentre le prime cinque per copie cartacee arrivavano al 32% delle vendite su carta. La scelta del supporto non è dunque un dettaglio neutro: carta e digitale replica hanno livelli di concentrazione diversi e possono mettere in evidenza testate

differenti.

Il diverso andamento di quotidiani nazionali e locali

Neppure la contrazione del mercato procede allo stesso modo per ogni categoria. Nel 2024 le testate classificate come prevalentemente nazionali hanno registrato una riduzione delle vendite del 7,3% rispetto all'anno precedente, contro il 3,1% di quelle maggiormente locali. Sul periodo 2020-2024, però, il calo è stato del 15,2% per le nazionali e del 25,9% per le locali.

Questi numeri descrivono l'andamento delle copie, non la qualità dell'informazione né la composizione dei rispettivi pubblici. Mostrano però che "quotidiani italiani" non identifica un insieme uniforme: testate nazionali e locali operano con funzioni informative e dinamiche distributive differenti.

Quali testate conviene confrontare

La combinazione dipende dalla domanda. Per seguire una vicenda politica nazionale è utile mettere a confronto almeno due generalisti con gerarchie e commenti differenti. Se la decisione ha rilevanti conseguenze finanziarie, si può aggiungere un quotidiano economico; se produce effetti territoriali, una testata locale consente di verificare ciò che il dibattito nazionale lascia sullo sfondo.

Il confronto diventa più solido quando riguarda articoli pubblicati nello stesso momento e distingue tre livelli: i fatti comuni, gli elementi presenti soltanto in una testata e le interpretazioni esplicite. Le divergenze sui fatti richiedono una verifica ulteriore; quelle nella selezione e nell'enfasi mostrano invece il diverso lavoro editoriale.

Scegliere un quotidiano significa scegliere una gerarchia, un ritmo e un punto di osservazione. Affidarsi a una sola testata rende quella prospettiva familiare, ma non completa. Una lettura più affidabile nasce quando le differenze tra i giornali diventano uno strumento di verifica: non adesione a una voce unica, bensì confronto tra prospettive capaci di illuminare parti diverse dello stesso avvenimento.

L’articolo Che differenza c’è tra i quotidiani italiani? Come cambiano linea editoriale e notizie proviene da Il Kim Magazine.

https://www.ilkim.it/che-differenza-c-e-tra-i-quotidiani-italiani-come-cambiano-linea-editoriale-e-notizie/

CVE-2026-19490: primi tentativi di sfruttamento sull’authentication bypass di Citrix NetScaler

Un exploit pubblico il 2 settembre, i primi tentativi osservati il giorno dopo: la falla che consente di aggirare l’autenticazione su NetScaler ADC e Gateway è passata in ventiquattro ore dalla teoria alla pratica. Citrix dichiara che non esistono workaround, e il solo aggiornamento non chiude la partita.

Da rischio teorico a sfruttamento

osservato

Il 19 agosto 2026 Citrix ha pubblicato il bollettino CTX696939, che documenta due vulnerabilità in NetScaler ADC e NetScaler Gateway. La più grave è la **CVE-2026-19490**, un *authentication bypass* con punteggio CVSS v4.0 pari a 9.3, classificato come CWE-288, ossia aggiramento dell'autenticazione attraverso un percorso o un canale alternativo. Il vettore descrive un attacco sfruttabile da remoto, a bassa complessità, senza privilegi né interazione dell'utente, con impatto elevato su riservatezza, integrità e disponibilità del sistema colpito. La seconda, CVE-2026-19489 (CVSS 8.8), è un *memory overflow* che interessa unicamente le configurazioni con gruppi Large Scale NAT e SIP ALG abilitato, quindi con una superficie di esposizione molto più ristretta.

Per due settimane la CVE-2026-19490 è rimasta un rischio potenziale. Il quadro è cambiato quando, il 2 settembre, è stato creato su GitHub un repository con un *proof of concept* che richiama esplicitamente il bollettino. Il giorno successivo un sensore della società di *vulnerability intelligence* Previdian ha intercettato le prime richieste corrispondenti a quell'exploit. La scheda di tracciamento tenuta dalla società, aggiornata al 4 settembre, riporta dieci tentativi complessivi provenienti da sei indirizzi IP unici, geolocalizzati in Australia, Germania, Giappone e Stati Uniti, e attribuisce allo sfruttamento attivo un livello di confidenza medio, sostenuto da tre elementi: la telemetria propria, l'attestazione di terze parti e la disponibilità pubblica dell'exploit.

Ventiquattro ore tra la comparsa dell'exploit e i primi tentativi osservati: è questo il dato che dovrebbe orientare la priorità di intervento, più del punteggio CVSS.

Il numero degli indirizzi è cresciuto nel corso della rilevazione, ed è la ragione per cui in giro si leggono cifre

diverse: il 3 settembre Ryan Dewhurst, fondatore di Previdian, aveva parlato a BleepingComputer di tre indirizzi distinti tra Australia, Stati Uniti e Germania.

Due precisazioni sono doverose, e vengono dalla fonte stessa. Dewhurst ha sottolineato che si tratta di tentativi di sfruttamento, non della conferma di compromissioni riuscite su sistemi reali. E la rilevazione, come indica la stessa scheda, poggia su un solo sensore: è un segnale, non una misura della campagna.

Sul versante istituzionale, oltre al richiamo di Citrix di metà agosto, il Centre for Cybersecurity Belgium, che opera come centro nazionale di coordinamento NCC-BE, ha pubblicato il 4 settembre un avviso sui tentativi di sfruttamento della vulnerabilità, invitando gli amministratori a dare priorità all'aggiornamento di tutte le appliance vulnerabili.

Chi è vulnerabile, e perché questo bypass è difficile da vedere

La vulnerabilità interessa le appliance configurate come *virtual server* AAA oppure come Gateway nelle funzioni di SSL VPN, ICA Proxy, CVPN e RDP Proxy. Sono quindi coinvolti esattamente i dispositivi che governano l'accesso remoto: il perimetro logico dell'organizzazione.

Il dettaglio operativo che conta di più, e che nella copertura della notizia resta spesso implicito, riguarda i prerequisiti, che variano in funzione della *build* installata e della presenza di una *SAML action*:

- sulle versioni 14.1-43.56 e successive, e 13.1-61.28 e successive, lo sfruttamento richiede che sia configurata una *SAML action* oltre al Gateway o al *virtual server* AAA;
- sulle versioni 14.1-43.55 e precedenti, e 13.1-61.27 e

precedenti, è sufficiente la presenza di un Gateway o di un *virtual server* AAA, senza alcun requisito aggiuntivo;
- le installazioni 13.1 FIPS risultano interessate quando l'appliance è configurata come Gateway o *virtual server* AAA.

La conseguenza è controintuitiva ma decisiva: più il firmware è vecchio, più ampia è la condizione di vulnerabilità. Le organizzazioni che hanno rinviato gli aggiornamenti non sono soltanto in ritardo, sono esposte su una superficie più larga.

Una precisazione va fatta proprio sul ruolo del SAML, perché si presta a un equivoco che circola già. Che su alcune *build* serva una *SAML action* configurata non significa che la vulnerabilità consista nella falsificazione di asserzioni SAML o nella rottura della relativa catena di firme: la presenza del SAML è un prerequisito di configurazione, non necessariamente il meccanismo dell'attacco. Citrix non ha pubblicato dettagli implementativi, e gli analisti che hanno esaminato il caso si sono astenuti dal ricostruire il meccanismo in assenza di quei dettagli. Chi legge un *proof of concept* di terze parti farebbe bene ad applicare la stessa cautela alla descrizione che quel codice dà di sé.

Sul piano della rilevabilità, e non della gravità assoluta, un *authentication bypass* su un concentratore VPN pone un problema diverso e per certi versi peggiore di una esecuzione di codice remota: l'accesso iniziale non richiede di installare nulla. L'attaccante ottiene una sessione che, agli occhi dei sistemi a valle, è indistinguibile da quella di un dipendente legittimo. Nel momento dell'ingresso non c'è *dropper*, non c'è *beacon*, non c'è artefatto sul disco: c'è un utente che si autentica e comincia a muoversi. È il tipo di compromissione che sfugge alle logiche di *detection* basate sull'endpoint e che emerge, quando emerge, dall'analisi comportamentale delle sessioni.

Le versioni corrette e il nodo dei workaround

Le *build* che risolvono la vulnerabilità sono:

Ramo	Build corretta
NetScaler ADC e Gateway 14.1	14.1-73.32 e successive
NetScaler ADC e Gateway 13.1	13.1-63.21 e successive
14.1 FIPS	14.1-73.32 FIPS e successive
13.1 FIPS e NDcPP	13.1-37.277 e successive

Su questo punto serve una precisazione che cambia il modo di pianificare l'intervento: alla voce workaround e fattori di mitigazione, il bollettino Citrix riporta "None". Non esiste cioè un ripiego riconosciuto dal vendor, e l'unica risposta prevista è l'aggiornamento.

Circola, riportata da Help Net Security, la possibilità di una mitigazione basata su firme applicabile tramite NetScaler Console sui firmware successivi a 14.1-60.52 e a partire da 13.1-63.16, che dispongono della funzione Global Deny Lists. È una capacità del prodotto, non una misura che il bollettino avalli per questa vulnerabilità. Chi la adotta lo fa come tampone in attesa della finestra di manutenzione, e ha tutto l'interesse a verbalizzarla come tale: un ripiego documentato, con la sua motivazione e la sua scadenza, regge a un'ispezione; un ripiego non documentato no.

Sul fronte dell'esposizione, i conteggi Shadowserver riportati da BleepingComputer indicano oltre 22.000 appliance NetScaler ADC e quasi 1.700 istanze Gateway raggiungibili da internet a livello globale. Il numero va maneggiato con cautela: non è dato sapere quanti di quei sistemi siano honeypot, quanti abbiano una configurazione effettivamente vulnerabile e quanti siano già stati aggiornati. La vista collegata, che riguarda le istanze Gateway, è raggruppata per Paese, quindi chi vuole la fotografia italiana può leggerla direttamente lì, tenendo

presente che conta dispositivi esposti e non vulnerabili.

Un precedente di pochi mesi fa

Le ventiquattro ore tra exploit pubblico e primi tentativi non sono un’anomalia di questa vicenda, ed è quanto era già accaduto sullo stesso prodotto in primavera.

Il 23 marzo 2026 Citrix aveva pubblicato l’avviso per altre due falle NetScaler, le CVE-2026-3055 e CVE-2026-4368. La prima, anch’essa con punteggio 9.3, è finita nel catalogo delle vulnerabilità attivamente sfruttate della CISA entro una settimana dall’avviso e comunque entro la fine dello stesso mese, con un termine di pochi giorni imposto alle agenzie federali statunitensi per l’aggiornamento.

Il dato d’insieme è più eloquente del singolo caso. Secondo il conteggio di BleepingComputer sul catalogo della CISA, da novembre 2021 sono ventitré le vulnerabilità Citrix classificate come sfruttate in rete, sei delle quali utilizzate anche da gruppi ransomware. Su questi apparati la sequenza avviso, analisi pubblica, sfruttamento non è un’eventualità remota, è lo schema ricorrente.

Proprio questo precedente suggerisce una considerazione, che è redazionale e non un dato. Il catalogo KEV della CISA viene spesso usato, anche fuori dagli Stati Uniti, come soglia implicita per far scattare un intervento urgente, ma è uno strumento rivolto alle agenzie federali statunitensi e non ha valore prescrittivo per un’organizzazione italiana. Usarlo come innesco è per di più una scelta fragile, perché l’inserimento nel catalogo certifica uno sfruttamento già in corso e arriva quindi, per definizione, dopo che la finestra di prevenzione si è chiusa. Il caso in esame lo conferma dal lato opposto: alla verifica del 5 settembre la CVE-2026-19490 non risulta nel catalogo, pur avendo punteggio 9.3, un exploit pubblico dal 2 settembre e tentativi osservati dal 3. Il segnale che conta non è l’ingresso in un elenco estero, è la

combinazione di quei tre elementi, che qui era completa da giorni.

Threat hunting: cosa cercare, e perché la patch non basta

Aggiornare è necessario, non sufficiente. La lezione della CitrixBleed del 2023 (CVE-2023-4966) è stata proprio questa: un *token* di sessione sottratto resta valido dopo l'aggiornamento e consente, come rilevò allora Tenable, di aggirare l'autenticazione anche dove è attiva l'autenticazione a più fattori. Fu lo stesso vendor a raccomandare, dopo l'*upgrade*, la rimozione delle sessioni attive e persistenti con i comandi kill aaa session -all, kill icaconnection -all, kill rdp connection -all, kill pcoipConnection -all e clear lb persistentSessions.

Il bollettino per la CVE-2026-19490 non contiene indicazioni analoghe, quindi quella che segue è una sequenza derivata da quel precedente, non una prescrizione del vendor per questa vulnerabilità.

1. **Inventario.** Individuare tutte le appliance con Gateway o *virtual server* AAA configurati, rilevando firmware e presenza di *SAML action*. NetScaler Console mette a disposizione una funzione di rilevamento per CVE con avvio manuale della scansione.
2. **Aggiornamento**, con priorità alle appliance esposte su internet e ai firmware più datati, che come visto sono vulnerabili in condizioni più ampie.
3. **Terminazione delle sessioni** attive e persistenti dopo l'aggiornamento, seguita dalla rotazione delle credenziali e dei segreti gestiti dalle appliance.
4. **Ricerca retrospettiva** nei log. Le evidenze pubbliche partono dal 3 settembre, ma il punto di ingresso esiste da prima del bollettino e nulla esclude che qualcuno lo conoscesse per conto proprio: il 19 agosto è quindi il

punto di partenza minimo ragionevole, non un confine. Da cercare: sessioni prive di un corrispondente evento di autenticazione, richieste SAML anomale, accessi da aree geografiche o intervalli di indirizzi inconsueti, modifiche di configurazione successive ad attività sospetta.

5. **Correlazione a valle**, perché il segnale più affidabile non è sull'appliance ma su ciò che è accaduto dopo: autenticazioni verso applicazioni interne provenienti da sessioni VPN senza storia, movimenti laterali, enumerazione di *directory*.

Il punto quattro è quello che distingue una risposta seria da una spunta su un foglio di calcolo. Se dall'analisi non emerge nulla, il risultato va comunque verbalizzato: in un'eventuale interlocuzione con l'autorità, l'assenza documentata di evidenze vale infinitamente più dell'assenza di documentazione.

Vale la pena aggiungere che, per le organizzazioni che ricadono nel perimetro NIS, l'esito di quella ricerca non resta un fatto interno. Se emergono evidenze di accesso non autorizzato scattano termini di notifica stretti verso il CSIRT Italia, e le fattispecie da notificare non sono le stesse per i soggetti essenziali e per quelli importanti: ne trattiamo nell'analisi dedicata agli incidenti significativi di base.

Il nodo strutturale

Ogni ciclo si ripete con la stessa meccanica: dispositivo di frontiera, vulnerabilità pre-autenticazione, PoC pubblico, sfruttamento nel giro di ore. I concentratori VPN e i gateway di accesso remoto restano il punto debole strutturale delle architetture aziendali perché concentrano una funzione critica in un software monolitico, esposto per definizione e aggiornabile solo con una finestra di manutenzione.

Le contromisure che riducono davvero l’impatto sono note e poco praticate: limitare l’interfaccia di gestione a intervalli di indirizzi amministrativi, separare il piano di gestione da quello di produzione, applicare *virtual patching* a monte con un *web application firewall*, segmentare ciò che sta oltre il gateway in modo che una sessione compromessa non si traduca automaticamente in accesso alla rete piatta. Nessuna di queste misure impedisce lo sfruttamento della CVE-2026-19490. Tutte riducono ciò che l’attaccante ottiene una volta dentro, che è la variabile su cui si gioca la differenza tra un incidente da notificare e una crisi da gestire.

https://www.ictsecuritymagazine.com/notizie/cve-2026-19490-citrix-netscaler-authentication-bypass/

Fuga cervelli da Italia, giovani laureati stranieri in arrivo in realtà compensano: i dati

(Adnkronos) – Tra il 2019 e il 2024, poco più di 78mila italiani tra i 25 e i 34 anni con una laurea si sono trasferiti all’estero. Nello stesso periodo, però, quasi 88mila giovani stranieri della stessa fascia d’età, con almeno una laurea triennale, hanno trasferito la loro residenza in Italia. I dati della Cgia.

–

multimedia/news-to-go

webinfo@adnkronos.com (Web Info)

The post Fuga cervelli da Italia, giovani laureati stranieri in arrivo in realtà compensano: i dati appeared first on Key4biz.

https://www.key4biz.it/fuga-cervelli-da-italia-giovani-laureati-stranieri-in-arrivo-in-realta-compensano-i-dati/586332/