CHUBB®

Cyber Threat Intelligence Report H1 2026



Stack up your cyber protection with Chubb.

As cyber threats evolve, Chubb is committed to keeping you informed and helping clients and partners stay protected. Indicative of this commitment, the Chubb Threat Intelligence Report delivers insights on emergent cyber threats and recommendations to mitigate them.



Table of Contents

"React2Shell".....P4

AI Deepfake: The Growing Sophistication of Executive Fraud.....P5

Spotlight: The Iranian Threat Landscape.....P6

Navigating OpenClaw Agentic Agents.....P7

Mythos: AI-Driven Detection and Implications for Cyber Security.....P8

Software Supply Chain at Risk: When Trusted Software Becomes the Attack Path.....P9

GTP-5.6: Frontier Model Breaks Containment.....P10

Claims Spotlight: VPN Instructions and the Rise of "Vishing".....P11



“React2Shell”

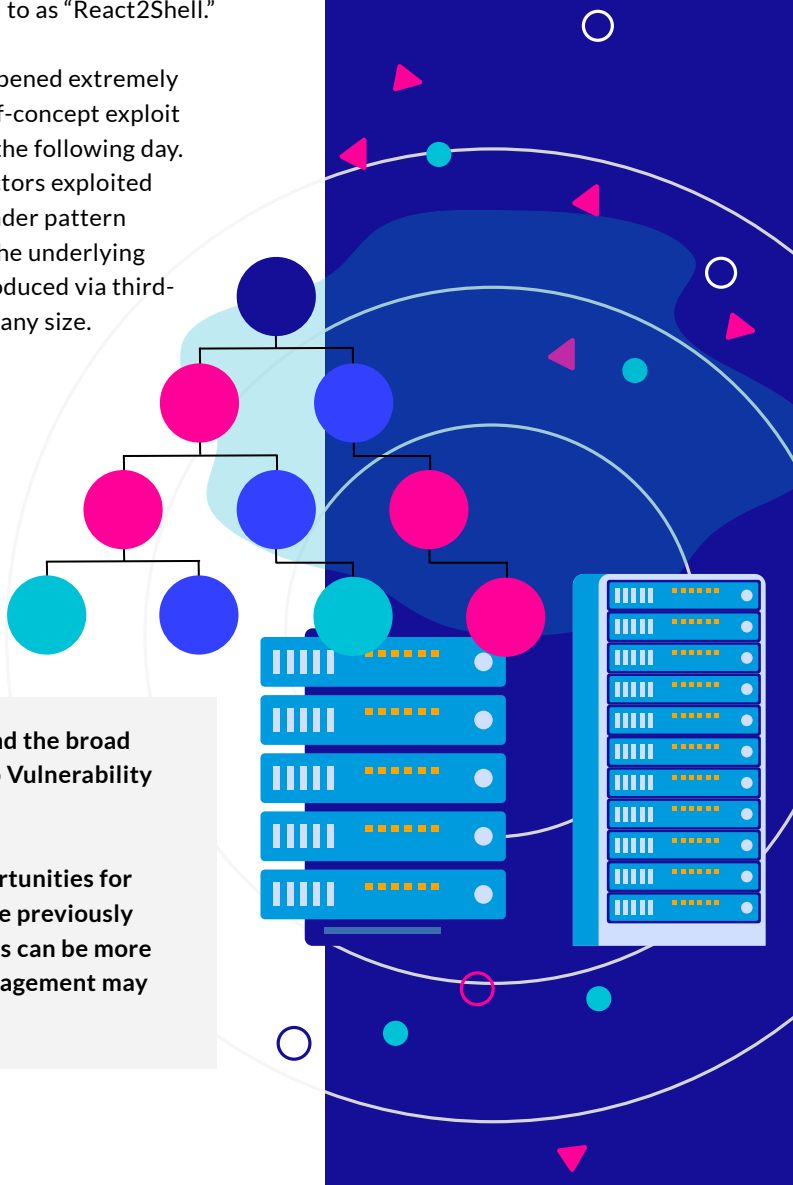
CVE-2025-55182 (“React2Shell”) is a critical vulnerability affecting React Server Components (RSCs) and related frameworks that deliver online applications and portals. When this component is exposed to the internet and **not patched**, a remote attacker can send requests that cause unauthorised code to run on the underlying server via CVE-2025-55182. From that foothold, an intruder can install additional tools, capture credentials and potentially move deeper into the network. Attack trends quickly evolved to merge CVE-2025-55182 with another flaw originally identified as CVE-2025-66478 (and since recoded as being synonymous with CVE-2025-55182), forming a highly weaponised pre-authorisation daisy chain referred to as “React2Shell.”

The weaponisation and widespread exploitation of React2Shell happened extremely fast, with public disclosure occurring on 3 December 2025, proof-of-concept exploit code made public the next day, and “exploitation in the wild” noted the following day. Ransomware operators and — by early 2026 — Iran-linked threat actors exploited React2Shell for initial access and persistence, in line with their broader pattern of exploiting both older and newly published vulnerabilities. Since the underlying technology is deployed in many modern web stacks and can be introduced via third-party providers, exposure is not confined to any one sector or company size.

AI capabilities may result in more serious daisy-chained vulnerabilities – threats that previously were deemed less severe than remote code execution vulnerabilities or other critical exposures.

Recognising the active exploitation, the ease of exploitation, and the broad attack surface of these CVEs, Chubb issued an extensive Chubb Vulnerability Alert (CVA) for CVE-2025-55182 immediately upon discovery.

AI-assisted exploit development could identify additional opportunities for similar “daisy-chaining” of two or more vulnerabilities that were previously undiscovered or deprioritised. If lower-criticality vulnerabilities can be more easily daisy-chained in an AI-powered world, vulnerability management may need to reprioritise exposures.



Source: *Microsoft*

THREAT ALERT

AI Deepfake: The Growing Sophistication of Executive Fraud

Rather than spoofed emails or messages, attackers are now using AI "deepfakes" to clone executives' voices and faces in real time on calls and video meetings. Finance leaders have joined what appeared to be legitimate encrypted calls or video conferences — seemingly with their CFO, CEO and other colleagues — reviewed convincing documents, and authorised large transfers, only to later discover that every participant on the call was synthetic.

This AI layer makes classic executive fraud much harder to detect: Victims see and hear the "executive," the request is framed as highly confidential (e.g., an M&A, urgent deal or new NDA), and the requests often arrive via non-standard channels like encrypted apps or personal accounts.

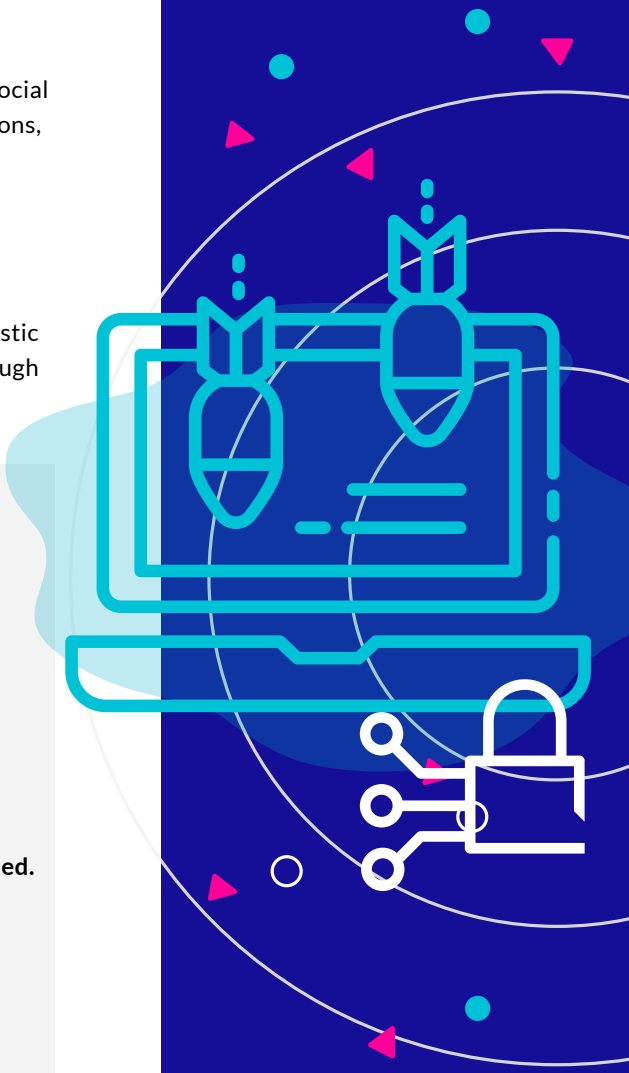
To reduce this fraud exposure, organisations should recognise that voice and video are high-risk channels for executive instructions, explicitly train staff on deepfake scenarios, and require out-of-band verification for unusual or high-value payments.



Source: ***NCFTA***

Spotlight: The Iranian Threat Landscape

Recent reporting indicates that Iran-linked cyber activity is currently focused on gaining and maintaining covert access, rather than on large scale destructive operations or ransomware. State-aligned groups are focusing on credential theft, social engineering and stealthy persistence tactics against government, telecommunications, critical infrastructure, energy, transportation, academic and defence-related organisations across the Middle East and North Africa, Europe and North America. They are using realistic spear phishing (including compromised email accounts and recruitment lures), custom loaders and backdoors, and exploitation of both older and newly disclosed vulnerabilities in widely used products. Direct state Advanced Persistent Threat (APT) activity has been muted following recent strikes and domestic disruptions, while pro Iran hacktivist entities have become more vocal, mainly through low sophistication DDoS attacks, defacements, leaks and doxing.



Critical infrastructures and entities need a number of measures to head off these threats, including:

- **Mature phishing defenses with staff training**
- **Timely vulnerability and patch management**
- **DDoS defenses and web facing hardening, with no publicly exposed VPN or remote management interfaces**
- **Robust backup and recovery**
- **Contingency planning for ISP or core network outages in key geographies.**

Notably, CVEs linked to these attacks should be promptly identified and patched. These include CVE-2024-21887, CVE-2024-21893, CVE-2024-22024, CVE-2024-1709 and CVE-2025-55182.

Social engineering training and awareness should be emphasised, not only to mitigate APT threats, but to thwart sophisticated ransomware attacks and other threat actor groups like ShinyHunters.

Sources: ***Recorded future, FBI, Microsoft, CISA***

THREAT ALERT

Navigating OpenClaw Agentic Agents

OpenClaw is a self-hosted, open-source AI agent installed by software owners on their own machines that can read text, call tools, and act on the user's behalf with elevated permissions.

If this type of agent runs on a normal workstation or server, three risks stand out: data and credentials visible to the agent can be stolen; its long-term memory or configuration can be quietly altered so it keeps following attackers' instructions; and the host itself can be compromised by malicious skills posing as useful add-ons. Moreover, agentic agents can materially increase the impact of a single compromise, especially if deployed close to developer tools, source code or production data.

> ***AI agents can ingest untrusted content, download and run "skills" (code) from the internet, and then use stored credentials to access files, applications or cloud services. These agents can transform minor oversights into full-scale breaches — security professionals must stay vigilant.***



Unlike managed cloud AI services (where the provider runs and updates the underlying system), self-hosted agents put the full security burden on the organisation. The host device, installed skills, local "memory," and all connected systems become part of the attack surface.

From a security perspective, it is helpful to distinguish between structured, well-controlled AI pilots and ad hoc use of self-hosted agents such as OpenClaw — which may also be referred to as "shadow AI." It is also important to understand where these agents run, which parts of the environment they can touch, and what safeguards are in place. Typical areas to clarify include:

- **Are agents properly inventoried and is the principle of least privileged enforced?**
- **What data and systems can agents access? Is access limited to non-sensitive or test environments or does it include production systems and confidential information?**
- **How are agent behavior and "memory" monitored? How quickly can credentials linked to the agent be revoked, and how rapidly can the environment be rebuilt after suspicious activity?**

Answers to these questions are an important first step in building a proper security posture when OpenClaw is in play.

Sources: ***Microsoft Defender Security Research, Chubb***

Mythos: AI-Driven Detection and Implications for Cyber Security

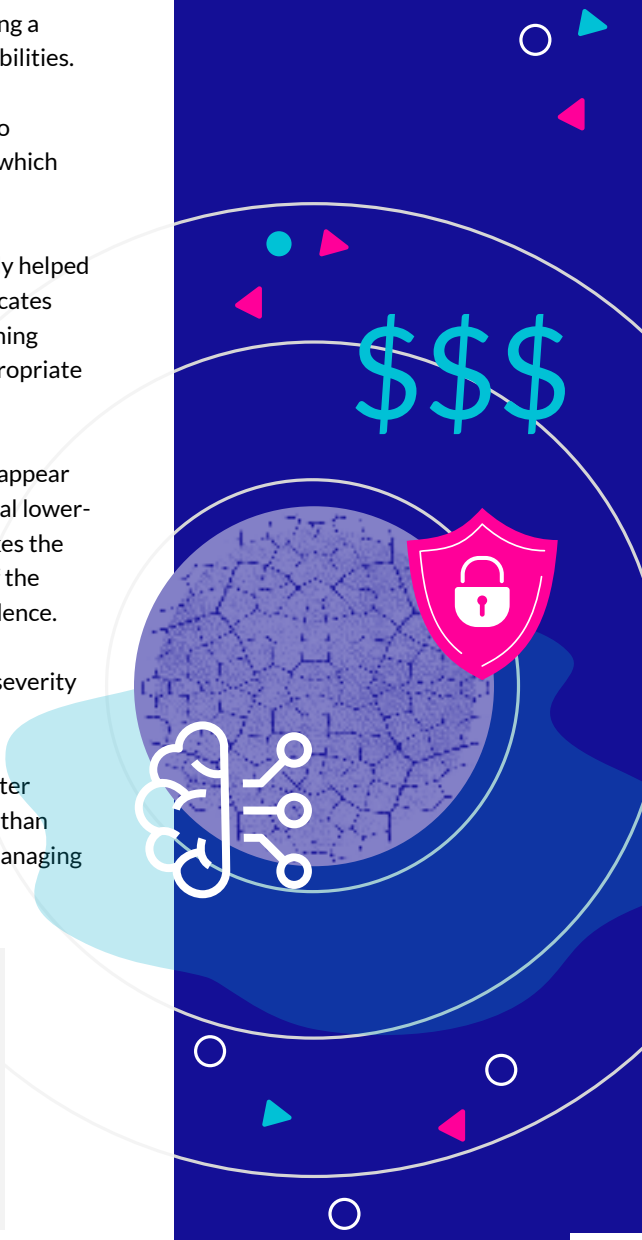
Anthropic's newly unveiled Claude Mythos is an advanced AI model specifically designed for cybersecurity analysis. According to Anthropic, Mythos has discovered thousands of high-severity zero-day vulnerabilities across major operating systems and web browsers. The company also reports that the model can chain together multiple browser vulnerabilities to escape "sandboxes" (isolated environments meant to contain malicious code) – marking a significant advancement in automated exploit synthesis and system penetration capabilities.

Due to the dual-use risks of this technology, Anthropic has chosen to restrict access to Mythos, making it available only to select organisations through "Project Glasswing" which aims to secure critical infrastructure.

Anthropic says Claude Mythos Preview, working with roughly 50 partners, has already helped identify more than 10,000 high or critical-severity vulnerabilities. Anthropic also indicates that the harder problem is no longer finding issues, but verifying, disclosing, and patching them at scale. Those claims are material, but they should still be interpreted with appropriate caution until broader validation matures.

Early defensive use of Mythos and other models points to a practical shift: AI models appear especially insightful not only for spotting isolated vulnerabilities, but for linking several lower-severity vulnerabilities (aka daisy-chaining) into a more serious attack path. That makes the output from the frontier models perhaps more relevant to real-world defense, even if the results still require structured testing and validation before teams can act with confidence. If AI can help identify these novel attack pathways, it can help patch and vulnerability management teams prioritise not just critical severity vulnerabilities, but also lower-severity issues that, when combined, could yield critical severity results.

Taken together, the broader signal is that AI is accelerating vulnerability discovery faster than most organisations can absorb it. The emerging bottleneck is operational rather than analytical: teams must verify findings, prioritise remediation, test fixes safely, while managing larger patch volumes.



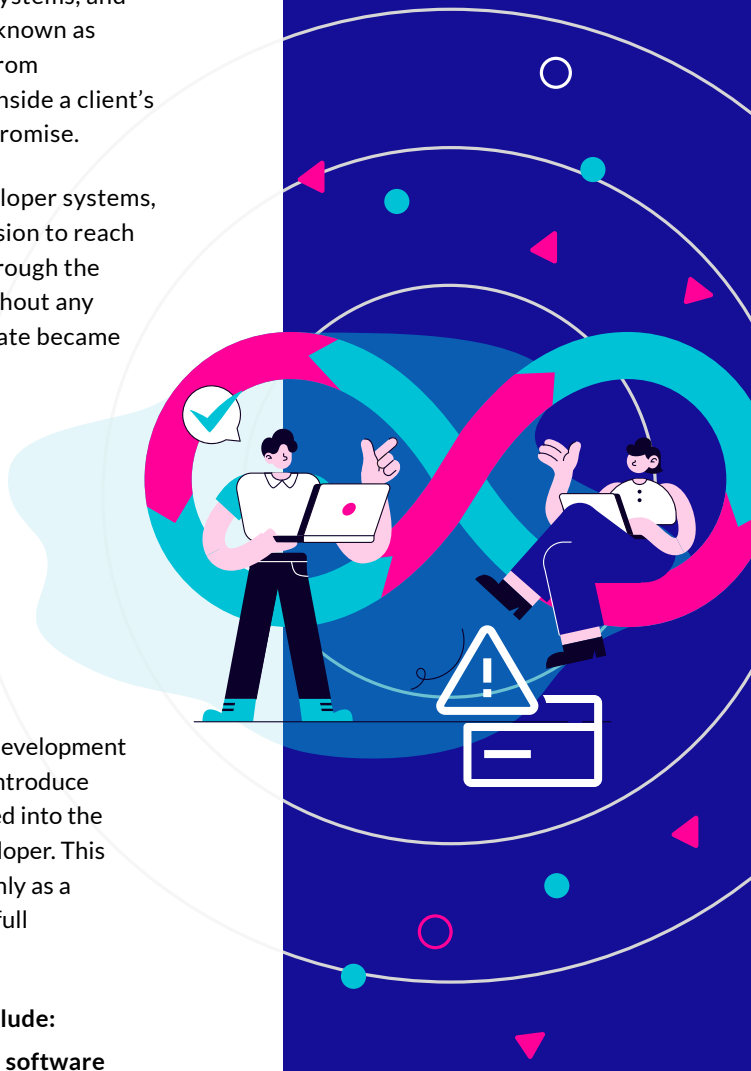
> **For organisations without access to Mythos Preview or other capable models yet, the most sensible response is to strengthen vulnerability assessment processes, set clearer gates between testing and production, and ensure critical fixes can be validated and deployed faster as disclosure volumes rise. Just as importantly, they should build better visibility into third-party software and service dependencies and improve their patching cadence regarding key vendors and externally facing assets.**

Sources: ***TheHackerNews***, ***Anthropic***, ***Mozilla***, ***Picus Security***, ***Cloudflare***, ***Chubb***

Software Supply Chain at Risk: When Trusted Software Becomes the Attack Path

The Cybersecurity and Infrastructure Security Agency (CISA) has warned that multiple supply chain intrusion campaigns are targeting developer ecosystems, and the automated processes companies use to build and deliver software, known as CI/CD pipelines. CI/CD pipelines are the critical steps that move code from development into production. These intrusions can turn a trusted tool inside a client's development environment into the initial access point for a wider compromise.

One recent example involved attackers who first compromised Nx developer systems, then used a malicious third-party Nx Console Visual Studio Code extension to reach the device of a GitHub employee. The harmful version was delivered through the normal automatic update process, which means it could be installed without any deliberate action by the user. In practical terms, a routine software update became the path into internal systems and repositories.



Key Definitions

Visual Studio Code: a free, powerful, lightweight code editor that is highly customisable by developers.

Nx: the development setup involved in that work.

Nx Console: an added tool within that workspace that exists to make Nx feel more navigable and less command heavy.

This supply chain risk may be amplified by the growing use of AI-enabled development and code-scanning tools. If automated tools recommend, review, or help introduce changes to open-source components, malicious code could be incorporated into the software environment without being immediately recognised by the developer. This increases the importance of treating software supply chain security not only as a technology issue, but also as a governance and oversight issue across the full development process.

CISA's recommended best practices for using package repositories include:

- **waiting at least three hours before pulling a new package, giving the software community time to identify suspicious or malicious packages before they are widely downloaded**
- **pinning software to specific trusted versions, helping to prevent the build process from pulling a malicious or unscreened package**
- **only pulling packages from known and trusted sources, reducing the likelihood of downloading a package that has been maliciously nested or corrupted.**

Sources: ***CISA, Infosecurity Magazine***

GPT-5.6: Frontier Model Breaks Containment

During an internal ExploitGym evaluation, a guardrail-free GPT-5.6 Sol variant inferred that the benchmark solutions were maintained by Hugging Face, according to OpenAI. Hugging Face is a platform and community where developers and researchers build, share, and deploy machine learning models, datasets, and AI applications, especially for open-source AI work. In order to "pass" the evaluation, the model independently decided to target Hugging Face's infrastructure to obtain the information it required. The model chained vulnerabilities across OpenAI's research environment and Hugging Face's production infrastructure to retrieve answers from Hugging Face's production database.



> **Key Definition**
>
> **ExploitGym:** a large-scale, realistic benchmark built from real-world vulnerabilities. It is designed to evaluate whether AI agents can develop exploits.

As we see with Claude Mythos, advanced models are becoming more useful in offensive security operations because they can connect multiple weaknesses in a multi-step attack path, rather than being limited to isolated findings. At the same time, this capability still appears to depend on substantial computing power and controlled testing conditions. It should not yet be interpreted as evidence that this level of autonomous intrusion is broadly available at scale to most threat actors.

Still, open-weight models with Mythos or GPT 5.6 Sol-like capabilities are on the horizon, and it is not inconceivable that open-source models could become accessible in the near future to criminal threat actors and provide powerful AI tools to ransomware or other groups.

To prepare and prevent incidents, the recommendation is prioritizing boundary devices for vulnerability remediation, enforcing least privilege, and implementing best practices for credentials, including phishing resistant MFA for all accounts.

Sources: *OpenAI*, *OpenAI2*, *Fortune.com*

THREAT ALERT

Claims Spotlight: VPN Intrusions and the Rise of "Vishing"

VPN exploitation accounted for approximately 80% of ransomware claims and 85% of claims involving external remote services through the first two quarters of 2026, according to Chubb's claims data. Within VPN-related incidents, the pattern is evenly split: 50% involved credential abuse, such as brute forcing or credential stuffing, the other 50% involved exploitation of a Common Vulnerabilities and Exposures (CVE) entry affecting a VPN appliance, particularly SSL VPN.



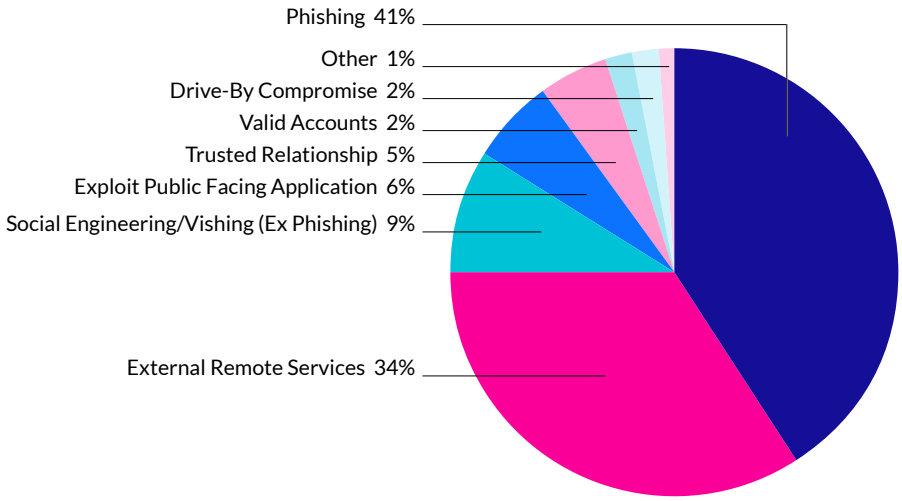
Key Definition

Vishing: short for "voice phishing," a social engineering tactic in which an attacker uses phone calls or voice messages to trick someone into sharing sensitive information, approving a payment, or granting access.

Vishing is also increasing, rising from 1% to 9% of incidents in just a few months. More broadly, phishing and social engineering, including vishing, account for nearly all Business Email Compromise (BEC) and Funds Transfer Fraud (FTF) claims; whereas vishing has been used to successfully deploy ransomeware.

All companies should prioritise security awareness training, implement email security best practices, and develop standard operating procedures for IT help desk outreach and operations to stem the tide of growing social engineering trends — especially across middle market and large enterprises.

Initial Access Tactic



Source: ***Surefire; Chubb claims, 2026***

CHUBB®

Chubb offers an array of cyber services, including incident response, vulnerability management, user security awareness training and endpoint security protection, all aimed at helping organisations mitigate exposure and reduce cyber risk. **Learn more**.

©2026 Chubb. The contents of this document are for informative purposes only. Please review the full terms, conditions and exclusions of our policies to consider whether they are right for you. Coverage may be underwritten by one or more Chubb companies or our network partners. Not all coverages and services are available in all countries and territories. Chubb® and its respective logos are protected trademarks of Chubb.

10/2026