FUJITSU

April 2026

Aura data breach: When one hour is all it takes

Mythos: AI and the compression of zero-day discovery

Developing securely: Practical steps to protect your software supply chain

When the guard opens the gate: How attackers are turning trusted security tools against organisations in Australia and New Zealand

Threat Intelligence Report

A monthly digest of cyber threat activities, insights, and strategies for enhanced cyber resilience

Fujitsu Cyber

Contents

This threat intelligence report has been developed using the insights from the various teams within Fujitsu Cyber. We report on the overarching trends we have recognised in the past few months, with a focus on current events and actionable steps.

At Fujitsu Cyber, we actively take these insights from what we observed and apply them to all the work we do, whether it be with our consulting engagements, our ongoing threat hunting programme, or our managed service client environments. Our constant learning across the business helps us to stay adaptable and on top of our security game, so that we can keep our client systems as safe as possible.

Aura data breach:

When one hour is all it takes

This article was written by:
Sonesh Seddiqi
Manager

In March 2026, identity provider Aura disclosed that it had suffered a data breach impacting approximately 900,000 records.

The incident was a result of a successful social engineering attack, where a threat actor gained unauthorised access to an employee account and used that access to extract data from internal systems.

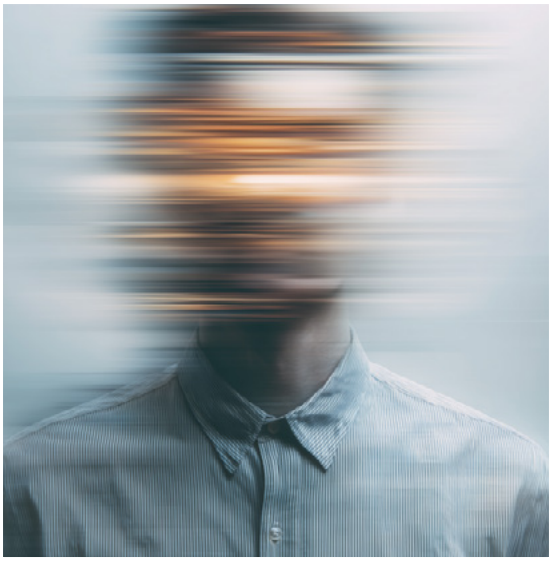
Aura is an identity protection and digital security company that provides services such as credit monitoring, identity theft protection, and online safety tolls. The platform is used by a large customer base, meaning any compromise involving user data can have broader downstream impacts beyond the organisation itself.

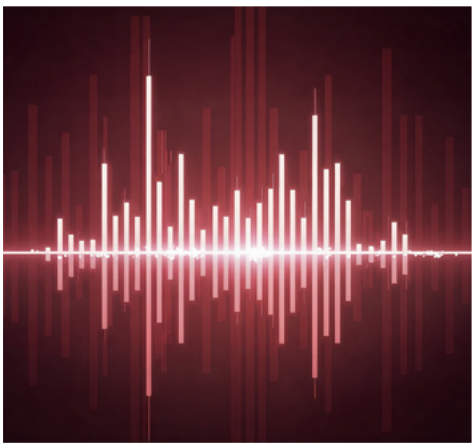
Initial reporting indicates that the attacker maintained access for roughly one hour, which was sufficient to exfiltrate a large volume of information. The group ShinyHunters has since claimed responsibility and alleged that the data has been released following failed ransom negotiations.

This incident highlights that even organisations focused on identity protection and fraud prevention remain vulnerable to attacks targeting the human layer, rather than technical vulnerabilities.

Whilst Aura confirmed that highly sensitive information such as passwords or financial data was not exposed, the dataset included names, email addresses, and contact information, with approximately 35,000 records linked directly to Aura customers.

This type of data is often considered “low risk”, however in practice it forms the foundations of targeted phishing, impersonation, and social engineering campaigns.





The attack we keep seeing

There is something slightly uncomfortable about how familiar this incident feels.

Phishing through email, voice or otherwise continues to be one of the most effective initial access techniques. Despite years of awareness training, security tooling, and user education, attackers are still able to gain access in relatively simple ways.

If anything, this problem is becoming more difficult, not less.

With the rise of AI, attackers are now able to:

- **Generate highly convincing email and messages.**
- **Mimic tone, writing style, and internal communication patterns.**
- **Conduct real-time voice impersonation with increasing accuracy.**

In that context, solely relying on users to constantly detect and prevent these attacks is becoming less realistic.

This doesn’t diminish the value of awareness training, it remains important, however it is becoming increasingly clear that relying on user vigilance alone is not a dependable control.

Technical details

The initial access vector for the Aura breach was a voice phishing (vishing) attack, where the attacker impersonated a trusted entity and convinced an employee to provide access or perform an action that enabled account compromise.

Once authenticated, the attacker accessed marketing dataset associated with a previously acquired company, suggesting that legacy or non-core systems were within reach of the compromised account.

From there, the attacker was able to:

- Query and access the dataset.
- Extract a large volume of records.
- Exfiltrate the data within a relatively short timeframe.

The scale of data accessed in approximately one hour indicates:

- Board access permission associated with the compromised account.
- Limited restrictions on bulk data access.
- Detection and response mechanism were not immediate.

Importantly, this was not a technically complex attack. It relied on legitimate access obtained through social engineering, which is often harder to detect than traditional exploitation.

Risk and recommendations

Designing for when phishing succeeds

A more realistic approach to modern security is to assume that phishing will occasionally succeed. Rather than focusing solely on prevention, organisations should:

- Implement secondary verification processes for sensitive actions.
- Avoid single points of failure in user-driven access decisions.
- Introduce context-aware access controls (device, location, behaviour).

Access control and data segmentation

The volume of data accessed suggests that least privilege principles were not tightly enforced. Organisations should:

- Restrict access to large datasets based on clear business requirements.
- Segment legacy and acquired system data from core environments.
- Apply controls to detect or prevent bulk data extraction.

Monitoring and response

The one-hour access window highlights the importance of detection speed.

- Monitoring for abnormal data access patterns (e.g. large queries, exports).
- Alerting on unusual user behaviour.
- Implementing rapid response capabilities to terminate suspicious sessions.

Zero Trust and defence in depth

This incident highlights the need for Zero Trust principles, where access is continuously validated rather than assumed. A defence in depth approach should ensure that:

- A compromised account does not result in unrestricted access.
- Multiple control layers (identity, access, monitoring, data protection) reduce overall impact.

In conclusion

The Aura breach is not particularly complex, but it highlights the current threat landscape. It shows that social engineering remains one of the most effective attack vectors. Data that appears low-sensitivity can still create real risk at scale. More importantly, it underpins a shift in how these attacks are evolving. With AI improving the quality and believability of phishing attempts, the line between legitimate and malicious interactions is becoming increasingly blurred.

From a security perspective, this means accepting a difficult reality:

We are not moving towards a world where phishing disappears, we are moving towards one where it becomes harder to recognise. And that makes resilience, not just prevention, the priority.

References

[1] V. Constantinescu, "Aura data breach exposes 900,000 records after phishing attack," Bitdefender, Mar. 19, 2026. [Online]. Available https://www.bitdefender.com/en-us/blog/hotforsecurity/aura-data-breach

[2] A. Mashanienkova, "Is Aura legit? Expert review and verdict for 2026," Cybernews, Mar. 25, 2026. [Online]. Available: https://cybernews.com/security/aura-identity-protection-data-breach/

Article two:

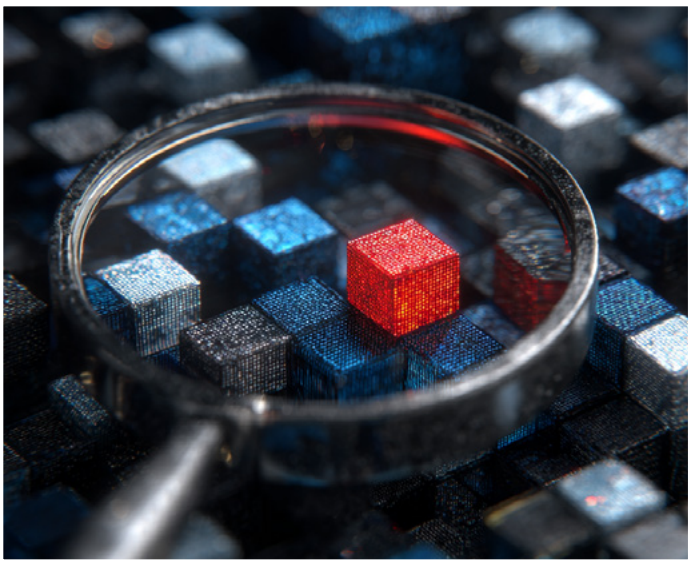
Mythos:

AI and the compression of zero-day discovery

This article was written by:
Adam McMullen
SOC Analyst

The use of artificial intelligence in cyber security is no longer emerging, it's established.

Across the current threat landscape, AI is being used to enhance phishing campaigns, automate elements of the attack chain, and increase the scale and efficiency of malicious operations. These developments are well understood and continue to evolve.



A more significant shift is now beginning to take shape.

Rather than improving how attacks are executed, AI is starting to influence how the vulnerabilities those attacks rely on are identified. This represents a transition from AI-assisted execution to AI-driven discovery.

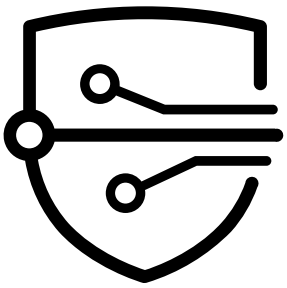
Emerging capability: Mythos

Recent reporting has highlighted Anthropic's unreleased frontier model, Claude Mythos, a system that has demonstrated the ability to identify large volumes of previously unknown vulnerabilities across widely deployed software environments [1], [3].

Unlike conventional tooling, Mythos has not been publicly released. Its deployment is intentionally restricted, reflecting the potential impact of its demonstrated capabilities.

In controlled environments, the model has reportedly:

- **Identified previously unknown vulnerabilities at scale.**
- **Analysed complex systems with minimal prompting.**
- **Generated exploit paths and linked smaller weaknesses into viable attack chains.**



In addition to identifying vulnerabilities, the model has demonstrated the ability to correlate and chain weaknesses into viable exploit paths, reducing the effort traditionally required to move from discovery to exploitation [2], [3].

This positions Mythos as something distinct from traditional security tooling. Rather than operating within predefined rules or signatures, it demonstrates the ability to explore systems in a way that is comparable to a human vulnerability researcher, but at significantly greater speed and scale.

Industry response: Project Glasswing

In response to this capability, Anthropic launched Project Glasswing, a coordinated initiative involving major technology and cyber security organisations including Microsoft, Google, Apple, AWS, and CrowdStrike [1], [3].

Rather than releasing the model publicly, access is restricted to trusted partners, with the objective of identifying and remediating vulnerabilities in critical systems before such capabilities become widely accessible.

This is not a standard product rollout. It is a controlled deployment model designed to balance defensive benefit with clear dual-use risk.

The existence of a coordinated industry response to a single model highlights its significance and signals a broader shift toward AI-driven vulnerability discovery as a step-change in cyber security.

This concern is not limited to industry alone. Regulators, including the Australian Securities and Investment Commission (ASIC) and Australian Prudential Regulation Authority (APRA), are already monitoring the implications of models like Mythos, particularly in relation to financial system resilience and the potential for large-scale vulnerability discovery [4]. Internationally, bodies such as the Hong Kong Monetary Authority (HKMA) have also begun assessing the impact of advanced AI capabilities on cyber resilience, reinforcing that this is not a regional issue, but a global one.

This is reflective of a broader shift, AI-driven vulnerability discovery is no longer viewed purely as a technical challenge, but as a systemic risk requiring a proactive and coordinated response.

Demonstrated outcomes

While full technical disclosure remains limited, early reporting around Mythos provides several concrete examples that illustrate the scale of its capability.

These include:

Identification of long-standing vulnerabilities in BSD-based operating systems, including issues that had reportedly remained undetected for nearly three decades.

Discovery of previously unknown flaws in widely used media processing libraries, including FFmpeg.

Identification of Linux kernel weaknesses, including cases where multiple low-severity issues could be combined into viable exploit paths.

In addition to individual findings, the model has demonstrated the ability to:

Analyse large codebases and system interactions with minimal input.

Surface edge-case behaviour that would typically require targeted manual testing.

Correlate findings across systems to construct higher-impact attack scenarios.

The significance lies less in any individual finding, and more in the consistency and scale at which such findings can be produced [1], [3].

Case study: AI-assisted discovery at scale

Within the Glasswing initiative, Mythos has been applied to widely deployed software and systems.

The outcomes provide a practical indication of what this capability enables.

The model has identified large volumes of previously undiscovered vulnerabilities, including issues that had remained undetected in production environments for extended periods.

In several cases, the model extended beyond identification, demonstrating how vulnerabilities could be exploited and how multiple low-severity issues could be combined into higher-impact compromise scenarios.

These outcomes are consistent with established research into automated vulnerability discovery but represent an escalation in both speed and coverage.

Analysis

The capability demonstrated by Mythos does not introduce new vulnerabilities. Instead, it changes the conditions under which vulnerabilities are discovered.

This shift is characterised by:

Reduced time required to identify vulnerabilities.

Increased coverage across systems, codebases, and dependencies.

Lower barriers to advanced vulnerability discovery.

Greater ability to correlate issues into exploit paths.

Collectively, these changes move vulnerability discovery away from a model constrained by time and expertise towards one defined by scale.

Risk and impact

As discovery accelerates, the time between a vulnerability existing and being identified decreases.

This has several implications.

The window available for organisations to respond is reduced. At the same time, the volume of identified vulnerabilities increases, placing additional pressure on triage and remediation processes.

Traditional patch cycles may struggle to keep pace, particularly in complex environments. Detection-based approaches also remain limited, as they are inherently reactive and dependent on known indicators.

The result is a shift in emphasis: Security moves from detecting attacks to competing on discovery speed.

Defensive considerations

To adapt to this shift, organisations should prioritise internal vulnerability discovery, ensuring weaknesses are identified before they are exposed externally.

Assessment models should evolve from periodic testing to continuous analysis, supported by persistent monitoring of systems, configurations, and code.

Detection strategies should increasingly incorporate behavioural and anomaly-based approaches, while response time should be treated as a critical control.

Importantly, the foundations for this approach already exist within most environments.

Platforms such as Microsoft Sentinel and CrowdStrike Falcon provide extensive visibility across systems and endpoints. The challenge lies not in data availability, but in the ability to analyse that data to identify unknown risk.

Strategic outlook: Project Glasswing

Project Glasswing demonstrates how this capability can be applied in a controlled, defensive context.

Rather than introducing new standalone platforms, it builds on existing security ecosystems, integrating AI-assisted analysis into established workflows.

This includes:

Continuous identification of potential weaknesses.

AI-assisted analysis to surface non-obvious vulnerabilities.

Reduction in time between discovery and remediation.

Fujitsu's existing partnerships across platforms such as CrowdStrike and Microsoft provide strong coverage across endpoint and SIEM environments, positioning it well to support this transition.

The capability represented by Mythos does not alter the nature of the vulnerabilities, it alters how quickly they can be found.

As AI continues to evolve, the defining factor in cybersecurity resilience will not be whether vulnerabilities exist, but who identifies them first.

References

[1] Anthropic, "Project Glasswing," 2026. Available: https://www.anthropic.com/glasswing
[2] Business Insider, "Anthropic's Mythos AI model raises cybersecurity concerns," 2026. Available: https://www.businessinsider.com/anthropic-mythos-cybersecurity-concerns-what-smart-people-are-saying-ai-2026-4
[3] ZDNet, "Project Glasswing: Microsoft, Google, Apple, Anthropic," 2026. Available: https://www.zdnet.com/article/project-glasswing-microsoft-google-apple-anthropic/
[4] S. Murdoch and Y. Ngui, "ASIC, APRA Among regulators monitoring Anthropic's Mythos", iTnews, Apr. 21, 2026 [online]. Available: ASIC, APRA among regulators monitoring Anthropic's Mythos - iTnews

Article three:

Developing securely:

Practical steps to protect your software supply chain

This article was written by:
Jacob Woods
Software Engineering Manager

Supply chain security has been in the news again recently, with the past month seeing several high-profile supply chain attacks where legitimate open-source projects have been compromised by malicious actors.

What can development teams do to safeguard themselves against attacks like this, and how can they prove they're doing it well?

Trivy, an open-source vulnerability scanner, was compromised by a threat actor going by the name of TeamPCP. Malicious versions would scan for and exfiltrate .env files, credentials, keys, and configuration files. This led to 75 of the 76 version tags in the repository being redirected to malicious commits for approximately three hours. [1]

TeamPCP later compromised other projects such as LiteLLM, an open-source Python library used to interface with AI providers, pushing two new malicious versions to the project. This attack also harvested credentials and installed backdoors. Malicious versions were available for about 40 minutes. [2]

For development teams, these incidents should be extremely concerning. How can we take meaningful steps towards securing our software development life cycle against attacks like this?

It's also important that such controls are measurable and demonstrable. Customers care about the security of their software and the company providing it. Negative publicity around breaches causes major reputational damage, and in the wake of security incidents, questions around what could have been done to mitigate the risk are inevitable. Can we prove that the steps we're taking are meaningful and appropriate?

Drawing on established security frameworks, let's discuss some basic controls we can put in place to defend ourselves.

Security and the software development lifecycle

Let's briefly consider secure software development holistically.

Most software development life cycles (SDLC) like Agile or Waterfall don't explicitly address security. [3] There is a risk that security becomes a low priority afterthought or requires costly post-hoc refactoring or rearchitecting to achieve. Worst case, the fallout of a major security incident that could have been avoided will almost certainly outweigh the additional time spent on securing software early.

Security in software development should instead be "shifted left" and addressed as early as possible in the SDLC. This way, security flaws are more likely to be caught sooner during design or implementation. This is obviously preferable to discovering them once the software is running in production and exposing vulnerabilities to the outside world.

The need to prioritise security early is especially true for supply chain attacks, which developers are vulnerable to from the earliest stages of prototyping, when the first install commands are run for a new project or feature.

Implementing and assessing security controls

If we want to develop securely (and have evidence that we are doing so), it helps to have authoritative frameworks we can create and assess controls against.

The US National Institute of Standards and Technology (NIST) published the **Secure Software Development Framework (SSDF)**. [3] The SSDF is effectively a collection of high-level secure software development practices based on existing established standards.

Importantly, it is designed to be applicable and practical for teams and organisations of all scales. It also focuses on outcomes, rather than prescribing how to achieve them.

OWASP also has a framework in the form of the **Software Assurance Maturity Model (SAMM)**. [4] Whereas the SSDF concerns itself with specific outcomes, the SAMM assesses the degree of capability. Each security practice has three maturity levels, with the higher levels meeting more sophisticated and stricter success criteria. In this way, SAMM can provide a guide for organisations to improve their processes over time.

It's also not expected for organisations to reach maximum maturity across all security practices. Instead, it's expected that the target maturity level for any given practice is decided after considering which areas the organisation most stands to benefit from investing in.

The OWASP SAMM and NIST SSDF are completely compatible with one another. In fact, OWASP maintains an official mapping between the two frameworks created in collaboration with NIST. [5]

These are only two of many frameworks that exist. Others might be worth considering depending on your organisation's needs.

Practical controls

Below are some examples of what these frameworks recommend as potential steps that can help mitigate the risk of supply chain attacks and better secure software development lifecycles.

Pin exact versions and/or establish a minimum age for dependency versions to avoid malicious new versions of packages like what happened in the LiteLLM attack.

Keeping to specific versions protects against new, potentially malicious versions of software being introduced during build processes.

Age requirements work similarly to defend against new compromises. Package managers like uv for python support minimum age requirements for dependencies. While it's feasible for compromises to remain undetected for long periods, they are also often caught relatively quickly (malicious LiteLLM packages were only available for about 40 minutes), so an age requirement as low as a day is a meaningful defence.

This aligns with both SSDF and SAMM's emphasis on having consistent and repeatable build processes using tools hardened according to best practice.

Perform regular audits of dependencies. This is one way to find out if you do depend on a compromised version of a third-party component. Consider scheduled audits using tools that check public vulnerability datasets, like audit commands built-in to package managers.

SAMM's second level of maturity under Software Dependencies specifies regular and automatic reviews for known CVEs. SSDF requires identifying vulnerabilities on an ongoing basis as well.

Know what your dependencies are with a Software Bill of Materials (SBOM). This allows organisations to identify if they're impacted by newly discovered supply chain attacks and mitigate proactively.

This is part of the first level of maturity in SAMM's Software Dependency area and can make satisfying the requirements of many other parts of SAMM and SSDF easier – for instance, SSDF names SBOMs as one way to help identify vulnerabilities after release.

Use a pull-through cache for dependencies and images. Using cached versions of packages and images can protect against existing trusted versions of packages or images being overwritten with malicious copies like the Trivy compromise. Many pull-through cache solutions can enable stricter controls around package versions or security scanning as well.

SAMM recommends establishing a central repository of dependencies as part of the second level of maturity for software dependency management. SSDF gives a similar example as a way to help ensure dependencies are well-secured.

As SAMM highlights, the right selection of security measures depends on your organisation's specific context and priorities. It's also important to consider the target level of maturity across different areas, as greater value may come from investing in more advanced controls in some areas than in others.

Conclusion

The supply chain is one of many attack vectors that development teams need to worry about, and supply chain attacks are a case study in the value of considering security early and throughout the software development lifecycle. Using frameworks like SSDF and SAMM can help teams build more securely in ways that are both practical and demonstrable.

If you'd like to learn more about practically implementing secure software development, please reach out and ask about how Fujitsu can help you.

References

[1] L. Abrams, "Trivy vulnerability scanner breach pushed infostealer via GitHub Actions," Bleeping Computer, 21 March 2026. [Online]. Available: https://www.bleepingcomputer.com/news/security/trivy-vulnerability-scanner-breach-pushed-infostealer-via-github-actions/. [Accessed 20 March 2026].

[2] K. &. J. I. Dholakia, "Security Update: Suspected Supply Chain Incident," LiteLLM, 24 March 2026. [Online]. Available: https://docs.litellm.ai/blog/security-update-march-2026. [Accessed 20 April 2026].

[3] M. Souppaya, K. Scarfone and D. Dodson, "Secure Software Development Framework (SSDF) Version 1.1," February 2022. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-218.pdf [Accessed 20 April 2026].

[4] OWASP, "Software Assurance Maturity Model," OWASP, [Online]. Available: https://owaspsamm.org/. [Accessed 20 April 2026].

[5] OWASP SAMM Project Team, "Tackling App Security with SAMM-NIST SSDF Mapping," OWASP, 6 February 2023. [Online]. Available: https://owaspsamm.org/blog/2023/02/06/samm-ssdf-mapping/. [Accessed 20 April 2026].

Article four:

When the guard opens the gate:

How attackers are turning trusted security tools against organisations in Australia and New Zealand

This article was written by:
Dan Broad
Head of Managed Security Operations

Cyber security has entered a new phase. Attackers are no longer relying solely on loud, obvious malware to break into organisations. Increasingly, they are targeting the very tools designed to stop them.

Think of it like this, the security guard is still standing at the front door, alert and capable, but someone has stolen the guard's radio, copied the access card and is now using that trust to move through the building unnoticed.

That is the significance of recent research showing how attackers can abuse trusted security controls such as Microsoft Defender configurations, built-in Windows tools and administrative privileges to weaken detection and accelerate attacks.

This matters today because many organisations across Australia and New Zealand have already invested in modern security platforms. The next challenge is making sure those platforms cannot be turned against them.

Why this is important right now

As of today, cyber threats across the region are increasingly shaped by three realities:

1

Identity Is the new perimeter

Attackers know it is often easier to steal a password, hijack a session token or abuse admin access than to deploy custom malware. Once privileged access is obtained, trusted tools can be manipulated from inside the environment.

2

Security products are high-value targets

Any platform with deep system access, endpoint protection, identity tools, backup systems, firewalls, or SIEMs, becomes attractive to adversaries. If they can weaken those controls, they can operate longer and with less resistance.

3

Regulators expect more than tooling

Boards, executives and regulators are increasingly asking the same question: Do your controls actually work under attack conditions? Having a product deployed is no longer enough. Organisations need assurance, monitoring, governance and recovery readiness.

What the research shows

Recent security analysis highlighted how adversaries may exploit trusted Microsoft security components or Windows-native capabilities after gaining sufficient privileges. Microsoft Defender remains a strong and widely trusted security control. The lesson is broader than one vendor – any trusted tool can be targeted if governance and monitoring are weak.

Common abuse paths include:

Adding folders or files to antivirus exclusions.

Disabling protections through policy changes.

Using legitimate signed binaries to run malicious code.

Leveraging PowerShell or WMI for stealthy execution.

Using admin tools to avoid detection.

These are often referred to as Living Off the Land techniques, using what already exists in the environment instead of bringing in obvious malware.

Why attackers use these methods

Because they work.

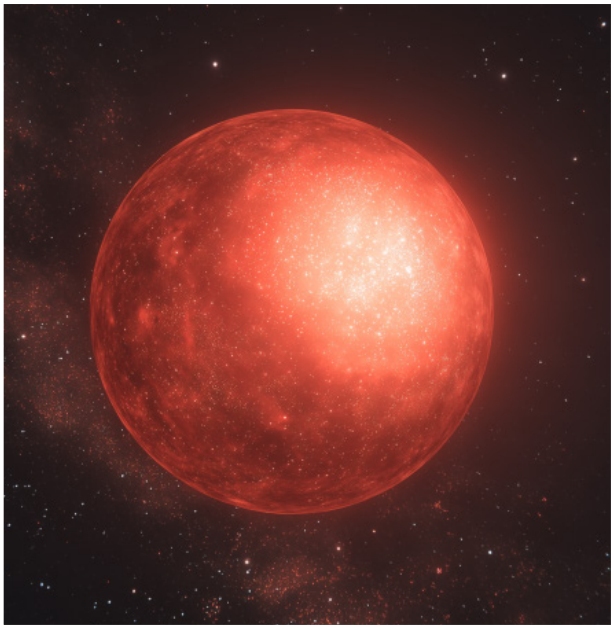
Traditional malware can trigger signatures, sandboxing, or behavioural alerts. Trusted tools often generate less suspicion, especially if security teams are focused only on malicious files rather than suspicious behaviour.

Attackers benefit by:

- **Blending into normal activity.**
- **Reducing alerts.**
- **Moving faster through the network.**
- **Delaying incident response.**
- **Increasing impact before detection.**

In ransomware incidents, even a short delay in detection can significantly increase operational and financial damage.

Example and proof of concept – RedSun



Publicly reported in April 2026, RedSun is a proof-of-concept privilege escalation technique that demonstrates how an attacker with an existing foothold on a Windows device could abuse Microsoft Defender's trusted remediation workflow to gain SYSTEM-level privileges. Rather than exploiting classic malware flaws, the PoC reportedly manipulates how Defender handles flagged files during remediation, redirecting a privileged file operation into a protected location and using that trusted action to elevate access. The significance is that the activity originates from a legitimate

security process, which can make detection more difficult. Reporting also noted similar tradecraft being observed in targeted intrusions, where attackers first gained low-level access, staged tools in common user folders, renamed binaries to avoid suspicion and then attempted escalation.

The lesson here is not that Defender is unsafe, as it remains a strong security control, but that any trusted tool can be abused if attackers gain access first. This reinforces the need for MFA, least privilege, Defender tamper protection, monitoring of exclusions or policy changes, detection of unusual SYSTEM activity, and blocking execution from user-writable directories.

ANZ impact

What this means for Australia

Australian organisations should be paying close attention, due to the heavy use of Microsoft ecosystems across government, enterprise and critical infrastructure.

Government

Agencies aligned to the ISM and PSPF already recognise the importance of privileged access control, logging and hardening. This threat reinforces why those controls matter.

Critical infrastructure

Energy, transport, ports, healthcare and utilities environments often rely on hybrid IT estates where Windows systems support business operations. If security controls are weakened early in an attack, disruption risk rises quickly.

Enterprise

Large cloud estates with broad administrator access remain attractive targets for identity-led compromise.

Small and medium business

Many SMEs rely on default security tooling without active monitoring of configuration changes. That gap can be exploited.

What this means for New Zealand

New Zealand organisations face similar risks, particularly where teams are lean and shared services are common.

Government

In the public sector, trusted environments and federated services can amplify the impact of a compromised privileged account.

Critical infrastructure, enterprise, and SMEs

Infrastructure and logistics operators, including transport, ports, utilities and exporters, remain attractive targets for disruption and espionage. Mid-market organisations often have solid tooling but limited in-house detection capability, meaning misuse of trusted tools may go unnoticed without managed monitoring.

For cross-Tasman businesses, shared identity platforms and cloud environments between Australia and New Zealand can significantly expand the blast radius of a single compromise.

What organisations should do now

Review security tool governance
Know who can change security policies, exclusions and protections.

Reduce privileged access
Apply least privilege, separate admin accounts, and strengthen MFA.

Monitor trusted tool changes
Alert on:

- New exclusions
- Disabled protections
- Service stoppages
- Suspicious PowerShell activity
- Policy tampering

Enable protective features
Use tamper protection, centralised logging and strong endpoint controls.

Test in real conditions
Run purple team exercises and incident simulations to confirm that controls effectively detect misuse.

Prepare recovery
Ensure the organisation can rapidly rebuild endpoints, identities and core services if trusted controls are compromised.

Conclusion

The conversation has changed. It is no longer enough to ask: Do we have the right tools? The better question is: Can we detect when trusted tools are being abused?

For organisations across Australia and New Zealand, that shift matters now. Those that adapt early will be better placed to withstand ransomware, insider misuse, identity compromise and the next generation of low-noise attacks.

The security guard is still valuable. But today, you also need cameras watching the guard room.

References

[1] Cyderes. "Redsun Zero-Day." Cyderes. https://www.cyderes.com/howler-cell/redsun-zero-day.
[2] Vectra AI. "When the Defender Becomes the Door: BlueHammer, Redsun, and Undefend in the Wild." Vectra AI. https://www.vectra.ai/blog/when-the-defender-becomes-the-door-bluehammer-redsun-and-undefend-in-the-wild.
[3] Dark Reading. "Exploits Turn Windows Defender Into Attacker Tool." Dark Reading. https://www.darkreading.com/cyberattacks-data-breaches/exploits-turn-windows-defender-attacker-tool.
[4] Microsoft. "Prevent Changes to Security Settings with Tamper Protection." Microsoft Learn. https://learn.microsoft.com/en-us/defender-endpoint/prevent-changes-to-security-settings-with-tamper-protection
[5] Microsoft. "Microsoft Defender Antivirus Updates." Microsoft Learn. https://learn.microsoft.com/en-us/defender-endpoint/microsoft-defender-antivirus-updates.
[6] MITRE. "T1562: Impair Defenses." MITRE ATT&CK. https://attack.mitre.org/techniques/T1562/.
[7] MITRE. "T1218: System Binary Proxy Execution." MITRE ATT&CK. https://attack.mitre.org/techniques/T1218/.

FUJITSU

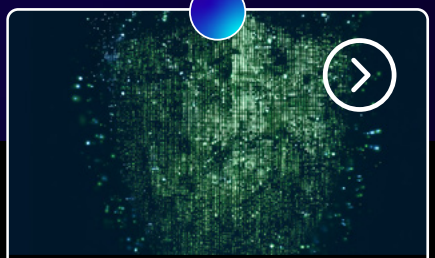
Fujitsu Cyber

Built to enhance cyber resilience

We are a Trans-Tasman team providing **end-to-end cyber security solutions designed to protect, enable, and transform organisations in Oceania.** We help you align with best practices, strengthen your defences, and ensure your systems are resilient and compliant. **Our cyber security services are structured around three core pillars:**

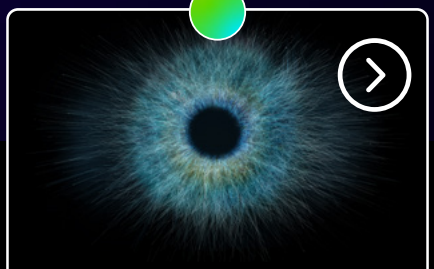
Advisory & Assurance

Delivering tailored consulting, strategic roadmaps, and hands-on support to help you identify risks, align with standards, and build resilience - empowering confident, secure business growth.



Technical Consulting

Uncover vulnerabilities and validate your defences through expert-led assessments and security testing. We deliver solutions aligned with business objectives while designing and implementing robust, future-ready security architectures.



Managed Security Operations

Providing 24/7 monitoring, proactive threat detection, and swift incident response to safeguard your organisation from evolving cyber threats.

Visit our website

Fujitsu Cyber draws on all parts of the business to identify key trends and changes with relevance to companies operating in New Zealand and Australia, both now and in the future. These threats are not solely technical. They can also arise from business operations, regional conditions in New Zealand and Australia, or global events that influence the cyber security environment in both countries.

Our research is the result of collaboration across the entire Australia and New Zealand team, including detection engineers, threat intelligence analysts, threat researchers, automation engineers, digital forensics and incident response specialists, as well as training and awareness professionals.

You may also be interested in our 'Best of 2025 Threat Intelligence Report' here

Authors:

Sonesh Seddiqi
Manager

Adam McMullen
SOC Analyst

Jacob Woods
Software Engineering Manager

Dan Broad
Head of Managed Security Operations

Curated by:
Thomas Hacker, Marco Pretorius, Hilary Bea

Compiled by:
Ed Goodacre
Digital Content Specialist

Contact us

Ready to strengthen you cyber resilience, get in touch