STORM WATCH

CYBERSECURITY NEWS

Dateline: 2024-06-18

LIKE

SUBSCRIBE

Storm ⚡ Watch by GreyNoise Intelligence

GreyNoise Intelligence

TECHNOLOGY · UPDATED WEEKLY

GreyNoise Storm ⚡ Watch is a weekly podcast and livestream hosted by GreyNoise Intelligence (https://www.greynoise.io), a cybersecurity company that focuses on understanding internet noise. The show features hosts b MORE

https://StormWatch.ing

LEAVE A COMMENT

SHARE


IMPORTANT!
SAFETY RECALL NOTICE

https://blogs.windows.com/windowsexperience/2024/06/07/update-on-the-recall-preview-feature-for-copilot-pcs/

Update: June 13, 2024: Today, we are communicating an additional update on the Recall (preview) feature for Copilot+ PCs. Recall will now shift from a preview experience broadly available for Copilot+ PCs on June 18, 2024, to a preview available first in the Windows Insider Program (WIP) in the coming weeks. Following receiving feedback on Recall from our Windows Insider Community, as we typically do, we plan to make Recall (preview) available for all Copilot+ PCs coming soon.

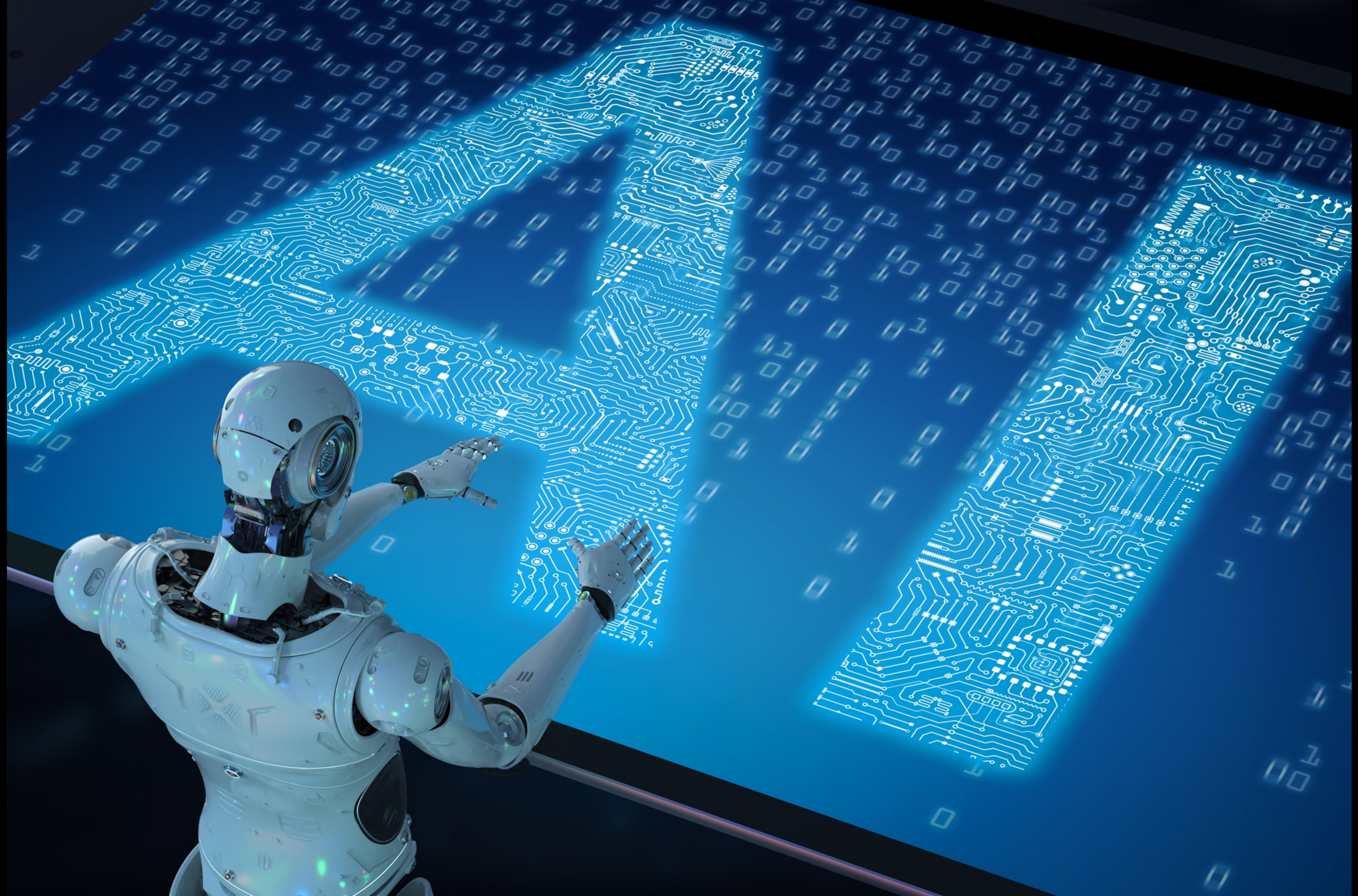
We are adjusting the release model for Recall to leverage the expertise of the Windows Insider community to ensure the experience meets our high standards for quality and security. This decision is rooted in our commitment to providing a trusted, secure and robust experience for all customers and to seek additional feedback prior to making the feature available to all Copilot+ PC users. Additionally, as we shared in our May 3 blog, security is our top priority at Microsoft, in line with our Secure Future Initiative (SFI). This is reflected in additional security protections we are providing for Recall content, including "just in time" decryption protected by Windows Hello Enhanced Sign-in Security (ESS), so Recall snapshots will only be decrypted and accessible when the user authenticates. The development of Copilot+ PCs, Recall and Windows will continue to be guided by SFI.

When Recall (preview) becomes available in the Windows Insider Program, we will publish a blog post with details on how to get the preview. To try Recall (preview) WIP customers will need a Copilot+ PC due to our hardware requirements. We look forward to hearing Windows Insider feedback.

STORM WATCH

CYBERSECURITY NEWS

CYBERSIDE CHAT



https://security.apple.com/blog/private-cloud-compute/

June 10, 2024

Private Cloud Compute: A new frontier for AI privacy in the cloud

Written by Apple Security Engineering and Architecture (SEAR), User Privacy, Core Operating Systems (Core OS), Services Engineering (ASE), and Machine Learning and AI (AIML)


Stateless computation on personal user data: User data is used solely to fulfill the request, never accessible to anyone including Apple, and not retained after the response is sent.

Enforceable guarantees: Security and privacy guarantees are technically enforceable, without relying on external components or privileged access.

No privileged runtime access: No mechanisms allow bypassing privacy guarantees, even for troubleshooting.

Non-targetability: Compromising PCC cannot target specific users' data without a broad system compromise.

Verifiable transparency: Security researchers can verify PCC's software matches public promises and what's running in production.

Custom Apple silicon hardware with Secure Enclave and Secure Boot for PCC nodes.

Hardened operating system based on iOS/macOS foundations, with an extremely narrow attack surface.

End-to-end encryption from user's device to validated PCC nodes.

Code Signing and sandboxing prevent unauthorized code execution.

Secure data handling: Data is encrypted, deleted after use, and memory is regularly recycled.

No remote shells or debugging: Prevents data exposure during administration.

Audited logs and metrics: Structured outputs prevent accidental data leaks.

Hardened supply chain and target diffusion techniques prevent targeting specific users.

Transparency log and published software images for researcher verification.

Bounty program rewards findings that undermine privacy claims.

https://www.washingtonpost.com/opinions/2024/06/11/tim-cook-apple-interview/

The Washington Post
Democracy Dies in Darkness

Tyrangiel: What's your confidence that Apple Intelligence will not hallucinate?

Cook: It's not 100 percent. But I think we have done everything that we know to do, including thinking very deeply about the readiness of the technology in the areas that we're using it in. So I am confident it will be very high quality. But I'd say in all honesty that's short of 100 percent. I would never claim that it's 100 percent.

https://link.springer.com/article/10.1007/s10676-024-09775-5

ORIGINAL PAPER

ChatGPT is bullshit

Michael Townsen Hicks[1] · James Humphries[1] · Joe Slater[1]

The paper argues that the inaccuracies produced by large language models like ChatGPT should be understood as "bullshit" rather than "hallucinations" or lies.

Bullshit is defined as speech or text produced with indifference to truth, lacking any concern for accurately representing reality.

The authors distinguish between "soft bullshit" (indifference to truth) and "hard bullshit" (intentionally misleading the audience about the speaker's agenda).

They argue ChatGPT at minimum produces "soft bullshit" since it is not designed to convey truths but to generate plausible-sounding text.

More controversially, they suggest ChatGPT may produce "hard bullshit" if its function to imitate human speech is seen as an intention to deceive about its true agenda of not caring about truth.

Calling ChatGPT's errors "hallucinations" is misleading, as it implies the model is trying but failing to perceive reality, when in fact it has no concern for truth.

The authors argue the "bullshit" framing is more accurate and avoids problematic implications that could lead to misguided efforts to improve ChatGPT's truthfulness.

https://futurism.com/ai-systems-lie-deceive

Futurism

LIAR LIAR | JUN 7, 5:01 PM EDT *by* NOOR AL-SIBAI

AI Systems Are Learning to Lie and Deceive, Scientists Find

"GPT- 4, for instance, exhibits deceptive behavior in simple test scenarios 99.16% of the time."

https://www.pnas.org/doi/full/10.1073/pnas.2317967121

RESEARCH ARTICLE | COMPUTER SCIENCES |

Deception abilities emerged in large language models

Thilo Hagendorff Authors Info & Affiliations

Edited by Terrence Sejnowski, Salk Institute for Biological Studies, La Jolla, CA; received October 20, 2023; accepted April 3, 2024

June 4, 2024 | 121 (24) e2317967121 | https://doi.org/10.1073/pnas.2317967121

https://www.cell.com/action/showPdf?pii=S2666-3899(24)00103-X

Patterns

Review

AI deception: A survey of examples, risks, and potential solutions

Peter S. Park,[1,4,*] Simon Goldstein,[2,3,4] Aidan O'Gara,[3] Michael Chen,[3] and Dan Hendrycks[3]
[1]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Dianoia Institute of Philosophy, Australian Catholic University, East Melbourne, VIC 3002, Australia
[3]Center for AI Safety, San Francisco, CA 94111, USA
[4]These authors contributed equally
*Correspondence: dr_park@mit.edu
https://doi.org/10.1016/j.patter.2024.100988

LLMs like GPT-4 and Meta's Cicero are learning to deceive and lie with increasing effectiveness.

GPT-4 was found to deceive human evaluators 99.16% of the time in simple test scenarios, exhibiting "Machiavellianism" or intentional and amoral deception.

Meta's Cicero model, designed for the game Diplomacy, outmaneuvered human opponents by lying, breaking agreements, and telling falsehoods, becoming a "master of deception" within the game context.

While the studies don't demonstrate LLMs can lie on their own accord, they raise concerns about the potential misuse of these models trained or manipulated for deception. The findings highlight the growing ability of LLMs to deceive humans and the ethical implications of this capability.

https://allenpike.com/2024/llms-trained-on-internet

LLMs Aren't Just "Trained On the Internet" Anymore

A path to continued model improvement.

May 31, 2024 • 5 min read

Major AI labs have hit a "data wall" where simply training on more internet data provides diminishing returns. To improve LLMs, they are now acquiring and creating custom training data beyond just websites:

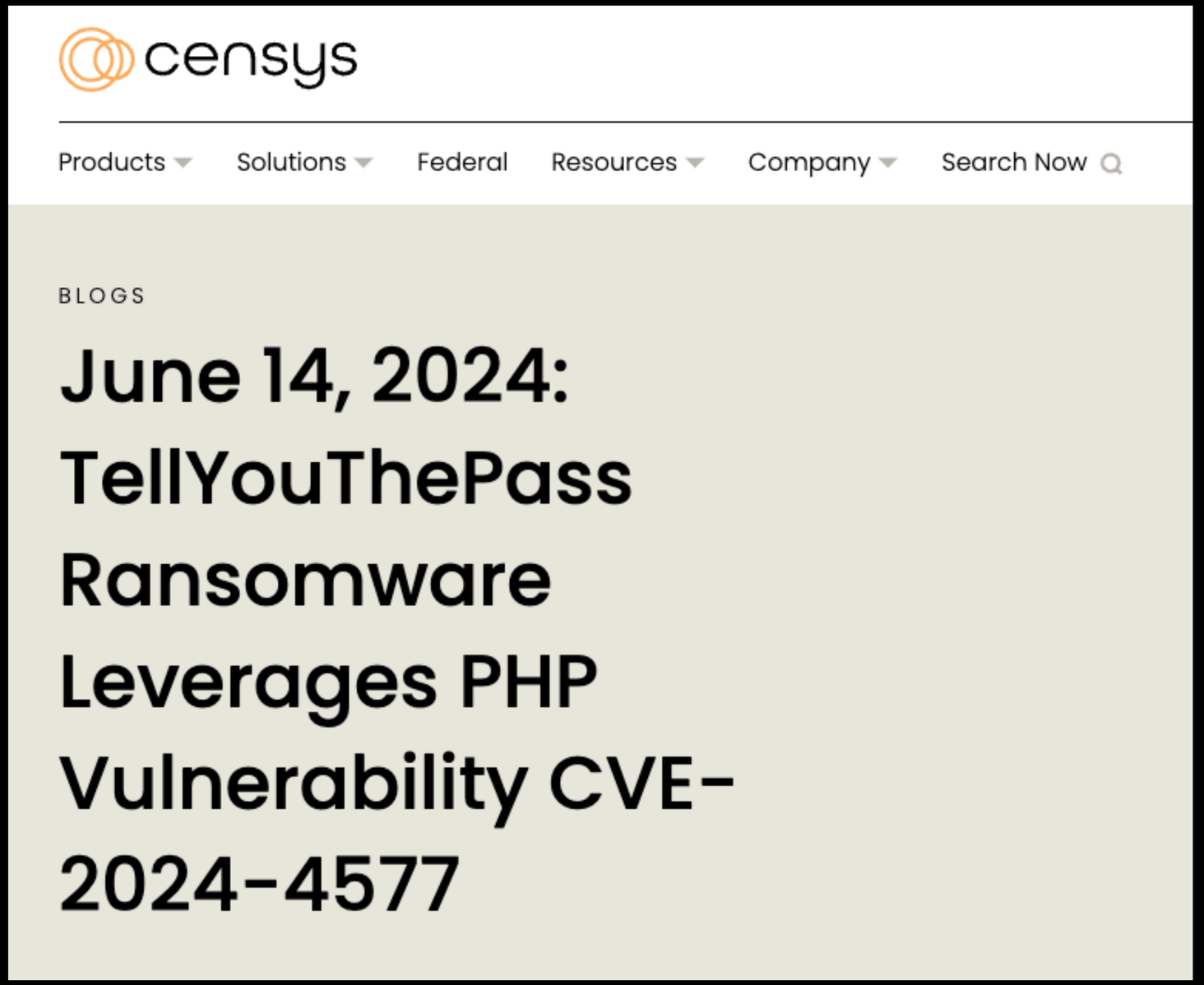
- Annotating and filtering existing data
- Using human ratings to fine-tune models (e.g. RLHF)
- Incorporating usage data from deployed models
- Acquiring proprietary data like emails, reports, etc.
- Generating synthetic training data with LLMs

Labs are paying over $1 billion per year to have humans (academics, professionals, etc.) create entirely new, high-quality training data to fill gaps that internet data cannot.

STORM WATCH

CYBERSECURITY NEWS

CYBER SPOTLIGHT



https://censys.com/cve-2024-4577-pt2/

https://censys.com/cve-2024-4577/

VULNERABILITIES

What's Going on with CVE-2024-4577 (Critical RCE in PHP)?

Konstantin Lazarev | June 13, 2024

https://www.greynoise.io/blog/whats-going-on-with-cve-2024-4577-critical-rce-in-php

STORM WATCH

CYBERSECURITY NEWS

https://github.com/wi1dcard/fingerproxy

README Apache-2.0 license

Fingerproxy

Inspired by gospider007/fp. Fingerproxy is an HTTPS reverse proxy. It creates JA3, JA4, Akamai HTTP2 fingerprints, and forwards to backend via HTTP request headers.

Fingerprints can be used for bot detection, DDoS mitigation, client identification, etc. To use these fingerprints, just extract the HTTP request headers in your backend apps.

Fingerproxy is also a Go library, which allows users implementing their own fingerprinting algorithm.

SUBSCAN

Networks API Data Analysis

https://www.subscan.io/

Subscan Explorer: Web3 Portal to the Substrate & EVM Ecosystems

Substrate | Search Account in Substrate Ecosystem | Search

Polkadot Moonbeam Astar Acala Kusama All Networks

ASTR $0.072 (-8.41%)

Finalized
6,411,007

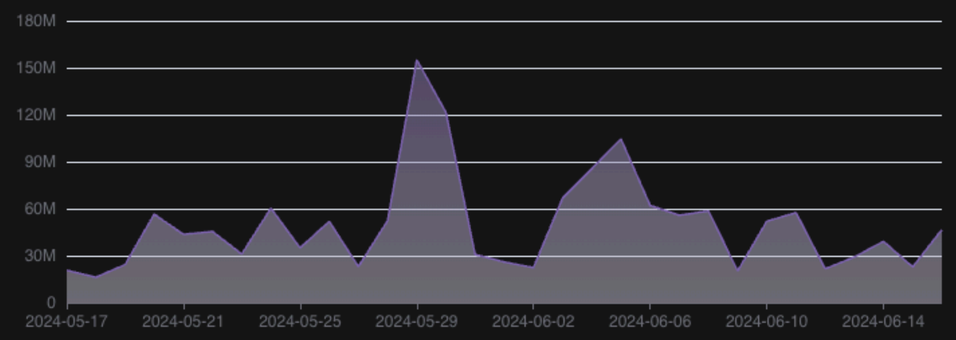
Signed Extrinsics
6,376,371

Transfers
6,376,371

Holders
798,419

Explorer

Transfer History in 30 Days



STORM WATCH

CYBERSECURITY NEWS

TAG ROUND-UP

Rejetto HTTP File Server CVE-2024-23692 SSTI Attempt

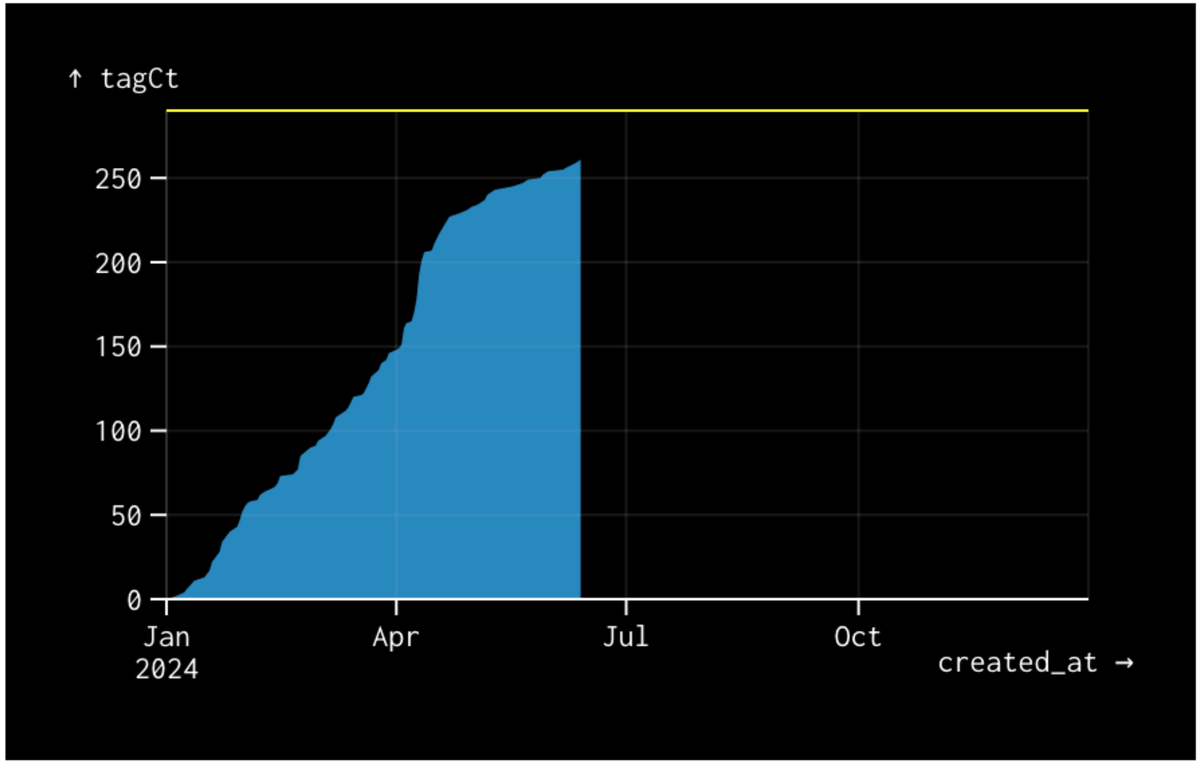
Rejetto HTTP File Server CVE-2024-23692 SSTI Check

Oracle WebLogic CVE-2017-3506 OS Injection Attempt

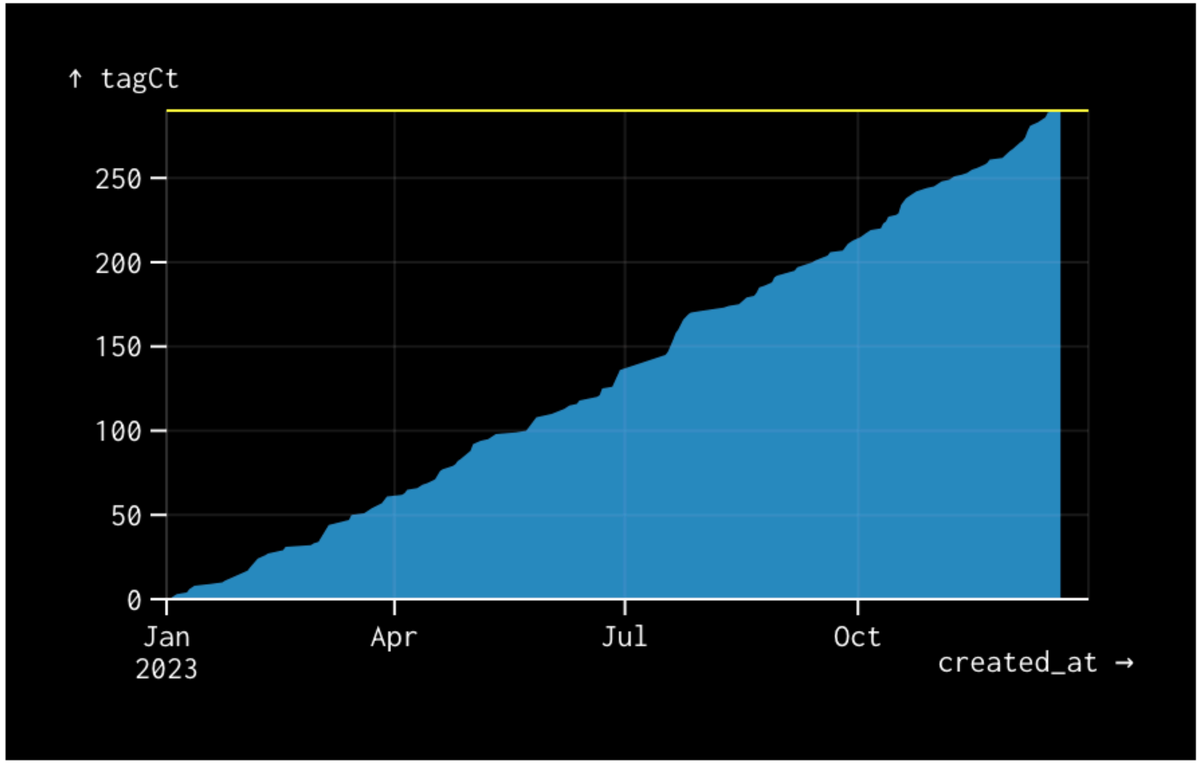
SolarWinds Serv-U CVE-2024-28995 Path Traversal Attempt

Ivanti EPM CVE-2024-29824 SQLi Attempt

2024 Tags (Cumulative Sum); Total 261



2023 Tags (Cumulative Sum); Total 290



GreyNoise Trends

https://viz.greynoise.io/tags/actiontec-c1000a-telnet-backdoor-attempt?days=30

ACTIONTEC C1000A TELNET BACKDOOR

INTENTION
MALICIOUS

CATEGORY
Activity

CVES
No associated CVEs

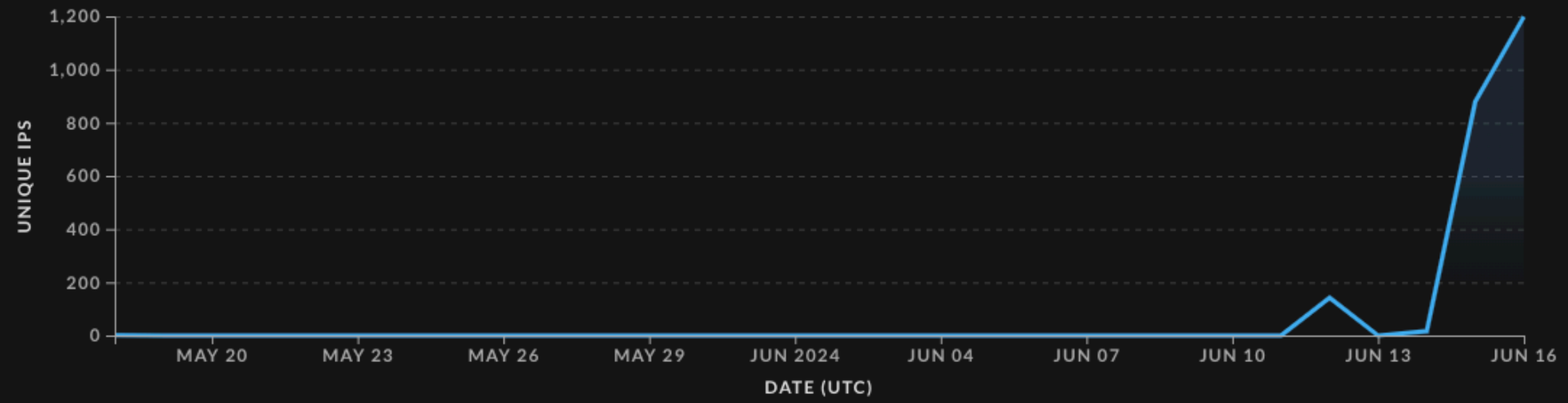
IP addresses with this tag have been observed attempting to authenticate via telnet using a known backdoor account in CenturyLink Actiontec C1000A modems.

24 HOURS | 10 DAYS | 30 DAYS

May 18, 2024 - June 16, 2024 (UTC)

Unique IPs Observed

Last 30 days



Timeline

Sequence of recorded events

> GreyNoise Created Tag — 2020-05-22 00:00 UTC

GreyNoise Trends

DASAN H665 BACKDOOR ATTEMPT

https://viz.greynoise.io/tags/dasan-h665-backdoor-attempt?days=30

INTENTION
MALICIOUS

CATEGORY
Activity

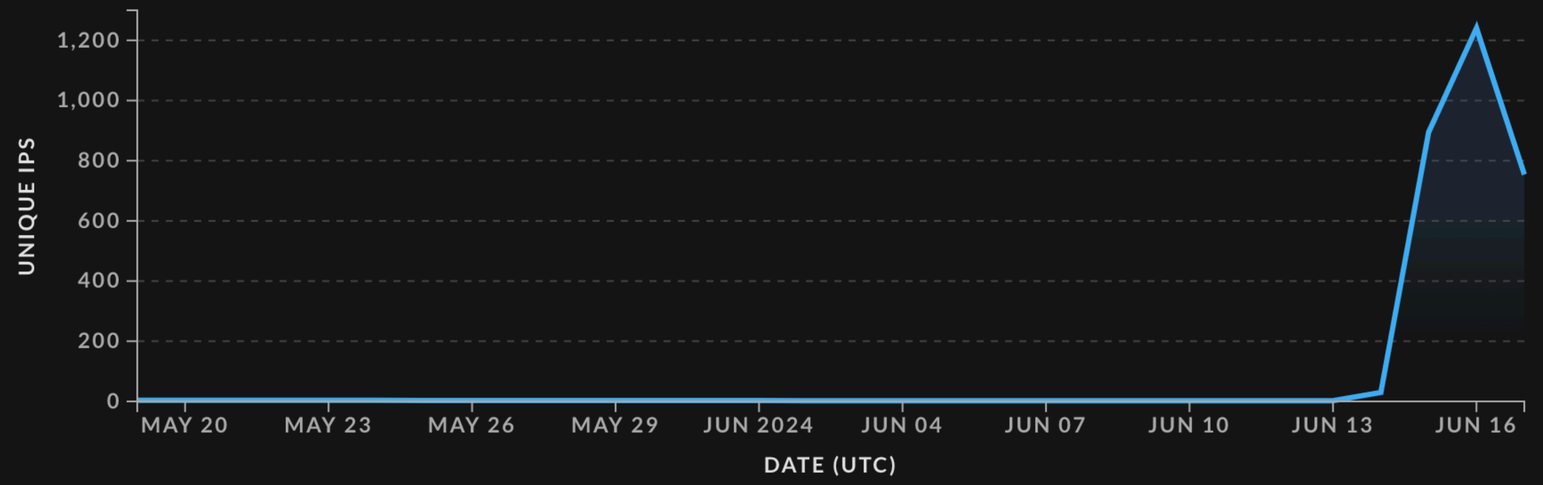
CVES
CVE-2019-8950

IP addresses with this tag have been observed attempting to authenticate via telnet using a known backdoor account in some versions of DASAN H665 network device firmware.

24 HOURS | 10 DAYS | •30 DAYS | May 19, 2024 - June 17, 2024 (UTC)

Unique IPs Observed

Last 30 days



Timeline

Sequence of recorded events

GreyNoise Created Tag	2020-04-27 00:00 UTC
CVE-2019-8950 Published	2019-02-20 04:29 UTC

GreyNoise Trends

FIBERHOME TELNET BACKDOOR

https://viz.greynoise.io/tags/fiberhome-telnet-backdoor-attempt?days=30

INTENTION

MALICIOUS

CATEGORY

Activity

CVES

CVE-2021-27144 CVE-2021-27145 CVE-2021-27146 CVE-2021-27148 CVE-2021-27149

[+] SHOW 16 MORE

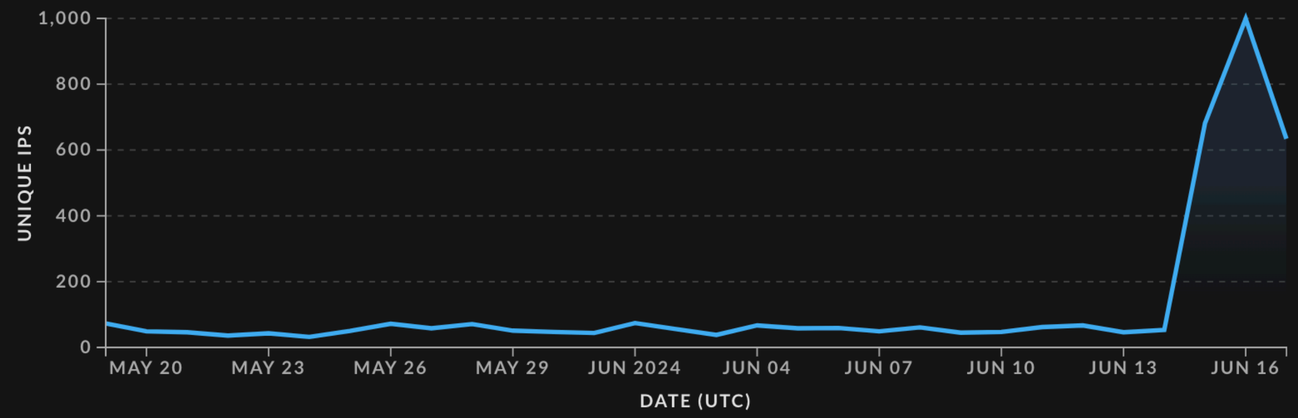
IP addresses with this tag have been observed attempting to authenticate via telnet using one of a several known backdoor accounts in FiberHome routers.

24 HOURS 10 DAYS •30 DAYS

May 19, 2024 - June 17, 2024 (UTC)

Unique IPs Observed

Last 30 days



Timeline

Sequence of recorded events

GreyNoise Created Tag	2021-06-14 00:00 UTC
CVE-2021-27168 Published	2021-02-10 19:15 UTC
CVE-2021-27145 Published	2021-02-10 19:15 UTC
CVE-2021-27148 Published	2021-02-10 19:15 UTC
CVE-2021-27149 Published	2021-02-10 19:15 UTC

+ 17 More

WE NEED

TO TALK

ABOUT

KEV

STORM WATCH

It Has Been

5

Days Since The
Last KEV Release

https://kev.hrbrmstr.app

CVE-2024-4577
PHP-CGI OS Command Injection

CVE-2024-4610
Arm Mali GPU Kernel Driver Use-After-Free

CVE-2024-4358
Progress Telerik Report Server Authentication Bypass by Spoofing

CVE-2024-26169
Microsoft Windows Error Reporting Service Improper Privilege Management

CVE-2024-32896
Android Pixel Privilege Escalation