The Residential Proxy Threat

Description: Worries of AI-power cyberattacks are spreading. Mythos "missed some" important vulnerabilities in Firefox. Every recent Patch Tuesday Nightmare Eclipse has struck. What now? Massive store of valid FortiGate VPN credentials found. F5 issues emergency updates to their NGINX-based server offerings. Introducing "AI Potpourri" - deeply altering an AI's personality. A close look at the explosion in malicious proxy networks. A Canadian judge okayed the illegal removal of such infections.

High quality (64 kbps) mp3 audio file URL: http://media.GRC.com/sn/SN-1084.mp3
Quarter size (16 kbps) mp3 audio file URL: http://media.GRC.com/sn/sn-1084-lq.mp3

SHOW TEASE: It's time for Security Now!. Steve Gibson is here. Lots to talk about. Some more vulnerabilities in Firefox. Steve has a theory about Microsoft's AI. F5 has some emergency updates. And we'll talk about malicious bad guys in your own network? The Malicious Proxy Networks. That's all coming up next on Security Now!.

Leo Laporte: This is Security Now! with Steve Gibson, Episode 1084, recorded Tuesday, June 23rd, 2026: "The Residential Proxy Threat."

It's time for Security Now!. Yay. You've been waiting all week. Tuesday's here, and so is Steve Gibson, the man of the day, of the hour, of the minute, of the second. It is time to talk about privacy, security, computers, and all of that. Hi, Steve. What are you doing there?

Steve Gibson: I've lost my fingerprints, apparently.

Leo: Well, that's not good. Or is it?

Steve: No. My iPads, I still have the iPads that are finger rather than camera.

Leo: The Mini is still that way, and I prefer that, especially [crosstalk].

Steve: Yeah, I like it a lot. And so I got one, I kept waiting to see if Apple was going to rev it, and it's like, okay, I can't - because the one I had, if you took it off of the plug, it was on life support, and it would just - well, actually what happened was some of the software that I want to use, like Claude's iOS app, and the dumb thing for my

thermostats, they don't run on the iOS that's so old, which is the last one that that Mini, that pad will operate on. So it's like, okay, I was kind of forced forward. Anyway, so because I've been terminating Cat5 cables and untwisted twisted pairs and straightening them out and pushing them through the Cat5 connector, I've kind of worn off my fingerprints in the last couple weeks. So anyway, that's what's going on.

We're going to talk about the surprisingly serious and growing threat from residential proxies. We've been touching on this recently because it's kind of been in the news. Well, there's essentially a recent explosion in the use of proxy networks because bad guys need to arrange to appear somewhere else. There's been a - we've heard me talking about it. So, like, why would you let Russians, people in Russian IP space, have any idea that you have an SSH server? Sure, you've got it protected with certificates and a really crazy password and multifactor authentication, but why should your SSH server even accept a connection from Russia? Are you planning to visit there?

So the point is that there's an increasing awareness of the value of geo-locking IP filtration. Which means that the bad guys have to work around that. How do they do that? They bounce their traffic off of unwitting people in the country that they want to target. Anyway, lots going on there we're going to talk about.

It was interesting to me, this awareness of the growing concern of AI's impact on cyberattacks is hitting the mainstream press. The Wall Street Journal had an article that I want to share parts of and interject some thoughts into. Also it looks like maybe Mythos missed some important vulnerabilities in Firefox, which is interesting because, what? Also every recent Patch Tuesday has had an attack by Microsoft's nemesis, Nightmare Eclipse. And that happened again. So we'll take a look at that.

Leo: He is really a thorn in the side of Microsoft.

Steve: Oh, my god. It's like, what? Why doesn't someone at Microsoft just say, okay, what do you want? Like we're sorry. Would you like a public apology? Would you like a Winnebago? You know, what can we do for you to solve this problem? Because, I mean, it's got to be expensive for them. You know, they're rattling their sabers and threatening. And that threat really did rub me the wrong way, as well as most of the rest of the security industry, because they're threatening someone over revealing something in software they have shipped containing about a bug that they put into the software. It's like, how is that this guy's fault for, like, not, you know, yes, he's not going by the unwritten rules. But they're unwritten.

And, you know, we just sort of made them up to help you, Microsoft, not for you to, like, you know, sue people over them. I mean, I'm sure that if you dissected their license agreement in the fine print it would say something like, you know, you agree by ever letting this software touch your computer that blah blah blah blah. So, you know, there's probably a technical basis for it. But that's not the way to solve the problem.

Anyway, we've got a really interesting massive store of known to be valid FortiGate VPN credentials found. You know, Fortinet is a leading supplier of border gateway devices. FortiGate is their VPN server. And someone's been getting into them and hoarding the credentials. We'll take a look at that.

Also F5, which purchased the company that was creating and publishing the NGINX web server, they just pushed some emergency updates which are important. I've decided to introduce what I will call the AI Potpourri section of the podcast, much as we've had, you know, Listener Feedback and so forth, which of course we'll always have, because I wanted a place to talk about just random AI stuff that was of interest. In this case, it's a

listener of ours, and I may have referred to him, like last week or the week before, because it was interesting. He arranged to - Will is his first name. We'll get to him deeply altering his AI's personality, instructing it to basically act like Agent Smith from the Matrix.

I do remember that I talked about it last week. I have in detail what he told his AI, and then I asked him for some sample prompts so that we could all see how the AI replied. It's pretty fun, so I'm going to share that. And I have a feeling that a lot of our listeners are going to take this up. Then we're going to, as I said, take a close look at the explosion in malicious proxy networks. And then, just in the news as I was finishing this, a judge in Canada okayed something that I didn't think we would see. So lots to talk about, and of course we have a Picture of the Week, which will be fun. We've wandered far afield from anything having to do with security. Like, okay, the gate in the middle of the field, that was - and the sheep standing behind it patiently waiting for it to open. I guess we'd already sort of lost the thread there. But this one's fun. Anyone who's married will understand this.

Leo: I have a question for you before we get to the Picture of the Week and our first sponsor of the week. And I just wanted to run something by you. The Economist is reporting that in testimony to the Senate, the head of the NSA told Senator Mark Warner that one of the reasons Mythos and Fable were banned by the Trump administration is that...

Steve: Oh.

Leo: ...it had breached almost all of the NSA's classified systems within a few hours during a red team test. Now, I have no reason to think that Joshua Rudd, the head of the NSA and U.S. Cyber Command, was lying to Mark Warner, or that Mark Warner misrepresented it. Rudd, according to the Economist, told Warner that this tool broke into almost all of our classified systems, not in weeks but in hours. Except that that tells me the NSA's classified systems are buggy as hell.

Steve: Yeah. I'm, like, disturbed that...

Leo: To me it's not disturbing that Mythos did, that Mythos is capable of that. It's disturbing that they had such buggy classified systems.

Steve: Yes. And what we know is that we haven't actually yet seen evidence of superhuman performance from any of the AI.

Leo: No, in fact, that's what everybody's saying.

Steve: It's just endurance.

Leo: Right.

Steve: If you got really good hackers and told them we've got a million dollars here for you, you know, sufficiently motivate them, they can do what the AI could do. It's not like the AI can do stuff no one has ever seen before.

Leo: Yeah.

Steve: So that's really interesting.

Leo: Yeah. It would certainly explain the ban. And we had Alex Stamos last week on Intelligent Machines.

Steve: Except that they're not that far ahead of everybody else.

Leo: That's what Stamos has been saying. And Stamos, by the way, that letter, it's FreeFable.org, that open letter to the White House had signatories, a hundred signatories including Paul Vixie, one of the founders of the Internet.

Steve: Yeah.

Leo: I mean, just like really good cryptographers.

Steve: A real who's who.

Leo: It was a who's who. And one of the contentions was it's not doing anything that other models can't do.

Steve: Right.

Leo: In fact, OpenAI apparently has its next generation of ChatGPT ready to go, 5.6, that has - they call it ChatGPT Cyber, that does many of the same things. So...

Steve: Yeah, and I have a sense that, although Microsoft is not talking about that codename MDASH, which I wish they would just rename, from what we know of it, I think it beats all of this. I think it's...

Leo: It found a lot of those vulnerabilities from the giant Patch Tuesday last week.

Steve: I think what Microsoft has is already in another - in the next-generation class.

Leo: Yeah.

Steve: They've talked about letting other people, you know, experiment with it and turn it loose. So the point is, I mean, there is so much money and so much focus and so much promise that I don't think I've ever seen the human race more motivated about something. Really. Except maybe solving COVID, like we need to...

Leo: Right.

Steve: ...cure COVID quickly, get a vaccine. There was mobilization. But here it's like, wow.

Leo: It's going to be an interesting year; isn't it, Steve.

Steve: Oh, boy.

Leo: Aren't you glad you're watching Security Now!? And if you're not a subscriber to the show yet, you really ought to be. It's free. Just subscribe in your favorite podcast client. And if you want to support us, of course, join the club, TWiT.tv/clubtwit. Or you can pay for individual episodes in Apple's podcast app. I think Spotify also has that built in. We'd prefer, of course, you subscribe to the whole network and support everything we do. But I think more important, more and more important, the stuff we're doing here, not only Security Now!, but on Intelligent Machines and on the TWiT flagship show. This is going to be a very interesting 2026. Picture of the Week time, Mr. G. I have not looked. I am a husband. So go ahead.

Steve: I gave this picture the caption "Too much paint was being returned by well-intentioned husbands."

Leo: And it is, I see a True Value hardware store sign.

Steve: Yes.

Leo: This I know from actual experience. And I bet you do, too. Tell us what it says.

Steve: Somewhere in a place called Healy is a True Value hardware store. And they've got one of those signs where they use the plastic letters that they stick on in order to, you know, like Happy Father's Day or, you know, Yay, 250 Years for the U.S. Anyway, this one says "Men choosing paint must have note from wife."

Leo: I'm sure this is a joke. I'm sure they're not serious.

Steve: I think that's probably true. But it makes a point that all of us, all of we who have been married...

Leo: We all know.

Steve: Yes.

Leo: And it's true, you're lucky because your walls, your new walls are all painted white.

Steve: White. No Eggshell.

Leo: But as every husband knows...

Steve: No Sundance, yes.

Leo: White is not white. There is eggshell white, there's ecru, there's white with a little gray in it. I don't know what that's called. That's really funny.

Steve: [Crosstalk]. Yes.

Leo: I don't know if this happened to you, but whenever the contractors, and we've had many, come to our door, and I open it, they say, "Is Lisa here?" Same thing. Same reason.

Steve: Well, and our contractors just this morning had this experience. We had some previous cabinets that we wanted to be mounted in the garage. And I said, okay, put them in the back corner. And so they started to, you know, screw them into the back corner. Lorrie came out and said, oh, that's all wrong. It's like, oh, god, okay.

Leo: That's why they always say "Is Lisa here?"

Steve: "We've been saying it's going to be in the corner, honey." "No, no, I've got stuff that has to go the left of those, so they have to come out about a foot." And the workers kind of look at me, and they look at her, and I go, "She's the boss." So yes.

Leo: Lisa and I had this debate because they call me El Jefe. And I say, "And they call you La Patrona." I said, "Doesn't El Jefe outrank La Patrona?" She says, "No. La Patrona is the highest rank." That's the four-star general of the job. All right. What are we talking about today on Security Now!?

Steve: Okay. So I've been seeing more and more of the non-technical popular press beginning to rouse to the idea that AI is not just about chatting with an agreeable buddy, but that it is also enabling a new darker side. And this has, of course, been something we've already spent significant time looking at; but I wanted to share a sample of what the non-security-focused press is saying. One week ago, the publication The Atlantic - oh, I misremembered that. I thought it was The Wall Street Journal. But anyway, The

Atlantic, they ripped the Band-Aid off this with their headline "Assume You Will Be Hacked."

Leo: Wow.

Steve: Which was then followed - yeah.

Leo: That's terrifying.

Steve: Yeah, followed by the - yes. Like, hey, everybody. Just give up. Just step away from your phone or your computer. They followed with the tease: "AI is enabling a deluge of cyberattacks the likes of which we've never seen before." And of course apparently the NSA would agree with them in that position.

Leo: Yeah.

Steve: So the author of this wrote: "Late last month" - oh, and he's writing in the first person. He said: "Late last month I began to consider withdrawing some money from my savings account to buy gold." Whoa. He says: "It's the first time I've ever thought about panic buying. For all the firewalls and two-factor authentication codes, the safety of the Internet is starting to falter. Hackers are gaining the upper hand over organizations around the world - hospitals, energy grids, government agencies, and, yes, banks.

"As AI tools have become extremely good at writing code, they've also become extremely good at pulling off cyberattacks." He says: "Malware, after all, is still software. The result has been a change in the scale, speed, and sophistication of hacks that's difficult to overstate. Among its tens of thousands of clients, the cybersecurity firm Palo Alto Networks identified a fourfold increase in daily attacks from 2024 to 2025. Hackers are developing AI-enhanced computer viruses that adapt on the fly to avoid detection. They're automating cyberespionage campaigns on foreign governments. They're stealing data in minutes instead of hours." And Leo, none other than "Alex Stamos," he writes, "a former chief security officer of Yahoo and Facebook, told me: 'There's a crazy amount of offensive activity happening right now. Companies are getting hacked every single day.'"

Leo: We know that, yeah.

Steve: Yeah. A pause to note that, you know, everybody knows my position on this; right? My feeling is these are the chickens that are finally coming home to roost.

Leo: Yes.

Steve: Massively wealthy companies such as Microsoft and Cisco have been far more focused upon adding new features which obsolete their current products, rather than investing in the security of their current offerings. You know, it's like, "Oh, we're sorry, we're no longer offering support for the equipment we previously sold you and profited from. And just to verify how big a sucker you are, we suggest you do that again. Look at

this shiny new product that pretty much does the same job as the old one, which now sadly" - oh, boo hoo - "is no longer secure."

Right. I have very little sympathy for mega corporations who rely upon the charity of other researchers to find the flaws that should never have shipped in their products to begin with. They've all been quite willing to ship flawed products and just hope for the best. Since all of this Internet stuff began, the industry has limped along, existing in a state of precarious balance. And now AI is indeed threatening to rock that world by upsetting that status quo.

So The Atlantic continues, and oh, boy, are they singing my tune. They wrote: "If the NSA is perturbed by the rise in cyberattacks, which it apparently is, then surely my savings" - he's talking about in the bank - "are vulnerable. There could be any number of weaknesses in my bank's IT systems to directly hack. Or perhaps an AI-written phishing email targeted at an employee, personalized to sound like a family member or manager, could let hackers into the bank and empty my coffers. Even if the bank has great cybersecurity, an attack on another business - a medical clinic I visit, a car-rental company, a newsletter subscription - could steal my payment information and, potentially, much more. The attack angles are seemingly infinite. And no one is adequately prepared.

"The term 'software engineering' has always been an insult to the level of rigor demanded of mechanical, civic, and other engineers. Computer programs can be riddled with vulnerabilities and run just fine for years or decades, as much of the software underlying the web has done just that. Stamos, who is now the chief security officer at the AI-coding company Corridor, said: 'We've just been writing software in a totally slapdash and insecure way for decades now.' With some small high-stakes exceptions, such as software used on the International Space Station or nuclear submarines, code is written and deployed without much rigorous testing. If a bug is reported, okay, it gets patched.

"Such a relaxed security posture has been more or less fine because discovering vulnerabilities is difficult, and skilled hackers are few in number. Either nobody found the bugs, or nobody was able to exploit them. But traditional cybersecurity methods don't cut it anymore. Before, you might scramble for a week to patch a hole, Giovanni Vigna, a cybersecurity expert at UC Santa Barbara, told me. 'Now you could have hundreds of those every week.' Moody's Ratings has found that the time attackers take to exploit a publicly known vulnerability (the digital equivalent of a robber plotting how to get around a bank's guards and cameras after obtaining a key) fell from more than 700 days in 2020 to just 44 in 2025, faster than the average time cybersecurity teams take to patch the bug.

"Governments and major companies are on high alert for AI-enabled cyberwarfare. The wake-up call came this spring, with the announcement of two extremely advanced cyber models" - we all know where this is going; right? - "Claude Mythos Preview from Anthropic, and the analogous GPT-5.5-Cyber from OpenAI soon after. 'Many independent cybersecurity experts have told me,' he writes, 'that these models are as, or nearly as, skilled as elite human hackers,' which is why Anthropic and OpenAI did not release them publicly. Instead, the AI labs have granted a small number of partner organizations and government agencies exclusive access to the unrestricted versions of these cyber models in the hopes of shoring up their IT systems." And I guess now the NSA, we know, is among those. And, yeah, let's hope the NSA tightens things up.

"And this month," he writes, "Donald Trump signed an executive order to expedite just that." That is, the shoring up of IT within the government using these models. "Organizations can guard against the coming deluge of AI-enabled hacks, most notably by using AI to detect and resolve vulnerabilities before cybercriminals can exploit them.

Anthropic has itself used Claude Mythos Preview to find thousands of bugs in open-source software packages, many of which went undetected for years or decades, that undergird much of the Internet. Mozilla used Mythos to fix more than 400 bugs in the Firefox web browser in April, roughly 20 times more than it fixes in a typical month.

"And having an AI agent monitoring for intruders 24/7 could be far more effective than periodic cybersecurity audits. If you've been noticing more frequent updates in your web browser, work software, smartphone apps, it may well be because software companies are now using AI to scan their software for bugs. But, if anything, these efforts are late. Even though they're not as powerful as Mythos, plenty of free and open-source AI hacking tools are allowing criminals with little technical expertise to marshal the equivalent of an army of hackers at their fingertips. Tools from Google, Anthropic, and OpenAI have guardrails intended to prevent them from being used for hacks, but they're not perfect.

"All three companies have reported more and more sophisticated, hacking attempts using their AI models. When the courseware Canvas was hacked last month, upending classrooms in thousands of schools and universities worldwide, AI likely played a role. And the criminal group responsible, a notorious hacking ring called ShinyHunters, is known for using AI in all sorts of scams. Just weeks later, Google cybersecurity researchers reported that ShinyHunters had hacked into an Oracle HR system and may have stolen data from more than 100 organizations. Meanwhile, the Trump administration has forced Anthropic to revoke all public access to the latest version of Mythos, taking away perhaps the most powerful cyberdefense tool we have from both the government and private sector.

"That does not mean you should withdraw your life's savings and buy gold. But a tremendous amount of change needs to happen in a very short period of time; open-source AI models will soon catch up to Mythos and GPT-5.5. The Internet needs upgrades 'at a Y2K-like scale,' Raffi Krikorian, the chief technology officer at Mozilla, told me, referring to a widespread fear that computer programs interpreting the digits '00' to mean the year 1900, rather than 2000, would bring down the web." Of course Y2K.

"'But IT professionals spent years preparing for and ultimately avoiding a Y2K apocalypse,' he said. 'With AI, we have months. No one company or government can demand the requisite collective action rapidly enough to completely secure our digital infrastructure. Wendi Whitmore, the chief security intelligence officer at Palo Alto Networks, told me: 'There's no way organizations across the globe are going to patch everything that needs to occur within the next three to five months.'

"At the same time that bots are making hackers more capable, the technology is also making the web less robust to attacks. Coding agents, due to their propensity to hallucinate, frequently write insecure code; and humans, in the thrall of vibe-coding, usually don't take the time to verify it. Spotty AI code has, for instance, reportedly caused multiple outages in Amazon's e-commerce services. Meanwhile, the AI models being integrated across the web - into Amazon, Google, your bank's customer-service department, and more - are themselves new, untested, and vulnerable to all manner of creative attacks that allow hackers to request passwords and personal information.

"A few weeks ago, a group of cybercriminals basically just asked Meta's customer-service AI to give them access to some 30,000 Instagram accounts (including the Sephora corporate account and the defunct Obama White House account), and the AI obliged. 'Some of our internal backend checks failed in this instance, but it wasn't due to the AI agent itself, and we've addressed the underlying cause,' Andy Stone, a Meta spokesperson told me.

"The near future," he writes, "is very likely to involve more frequent, and more severe, outages and hacks just like those affecting Canvas, Meta, and Amazon. Vigna said: 'We will see more of these disruptions. I think it's inevitable in the short term.' Smaller but crucial companies and organizations that are not web-native - think power plants, municipal-government agencies, credit unions - are especially vulnerable. They may be running all sorts of clunky legacy code, and lack the IT capacity or the financial resources to make the necessary upgrades. In many cases, the person who wrote the bulk of an organization's software might be retired or dead.

"Take hospitals, many of which are already struggling to combat data breaches and ransomware attacks. Hospital IT systems are full of valuable health and financial data, and the incentive to pay a ransom is high when patients' lives are on the line. 'It's not a matter of will to increase cybersecurity for hospitals,' John Riggi, the national adviser for cybersecurity and risk for the American Hospital Association, told me. 'It's a matter of resources and capabilities.' AI, he said, will make everything worse. And the greater burden is always on the side of the defense: Missing just a single vulnerability can permit a catastrophic attack. An Anthropic spokesperson told me that 'hospitals, utilities, and smaller banks run on software built by others,' which Mythos is helping secure. 'Software upstream protects the organizations downstream that don't have the resources to staff their own security research team.'

"Mozilla's Krikorian said: 'A worst-case scenario over the next year or so might look like blackouts across the United States, telecommunications companies being hacked, or our banking systems dealing with people losing money left and right.' Every cybersecurity expert I," writes this person, "spoke with for this story concurred. The next few months, couple of years, or even longer is going to be rough. Whitmore said: 'I hope it's not a catastrophic outage, but I am concerned that 2026 really could be the year that we see some sort of attack like that become very successful.' Anthropic estimates that a major cyberattack on just one of its 200 or so partner organizations could affect at least 100 million people.

"Collective action aside, some precautions exist that individuals can take short of liquidating into gold. Many of them are basic." Okay. "Use a password manager that auto-generates long passwords, keeping software updated, restarting devices to wipe viruses from their short-term RAM. Be extra wary of all sorts of phishing texts and other low-level scams. And you might consider simplifying your digital life by switching to a Chromebook, certain tablets, or another gadget that is a 'thin client,' meaning that very little software and data are stored on the device."

And finally: "Even in the most catastrophic scenarios, perhaps we can ride out the AI hacks. No one knows just how many bugs are out there. If there's a limited pool of vulnerabilities online, things will settle down once they are all found, whether by hackers or security audits. But Stamos said it's also possible that every time the top AI models reach a new threshold of capabilities, they discover a new pool of still more complex hacks. And so the chaos begins anew."

So I think that the truest thing anyone can say at this point is that we have no idea what's going to happen. But I wanted to share that to remind everyone, that was in The Atlantic. Not some security magazine. Not some podcast, some security-focused podcast. Our listeners know that everything I just read we've all shared on this podcast. Right? There was nothing new here. Even the expectations and the sentiments and the understandings of the balance of good versus malicious. All of that is, like, yeah, we all know that. But this is now surfacing in stuff that other non-techies read. So this understanding of what's going on, this awareness, is now moving out into the mainstream.

So, you know, when a hurricane-scale storm is heading toward Florida, nobody knows exactly what the consequences will be because storms are chaotic and unpredictable. We just have a sense of, like, uh-oh, this is a big one. So what's happening with AI is analogous. We know without doubt that we have filled the world with insecure software. Stamos said the same thing. And we know that a large language model AI is able to find many of the problems that have remained hidden, in some cases for decades. Basically it's about economics; right? We're not spending the money because these are hard things to find. AI makes them easier to find.

So whose AI will be the first to discover a previously unknown problem? The good guys or the bad guys? And even if it's the good guys, remember, and this is one thing that this guy did not note that we know about here on the podcast, the reality is that discovering the problem is, as formally phrased in the fields of logic or philosophy, it's necessary but not sufficient. You cannot fix a problem you don't know about. So knowing about it is necessary. But the goal is to remove the existing defective software from the field, and only doing that is sufficient.

As I've noted before, Patch Tuesday is no longer a Microsoft-only occurrence; right? Other companies have developed automatic software deployment systems. But the vast majority of buggy software that's currently deployed in the field is never updated. So even if publishers obtain new and improved AI-enhanced, AI-debugged software, the "sufficient" portion of the equation will remain unmet. Necessary but not sufficient, if you fix the bugs, but don't get them out into the field. I've noted how fortunate it is that all modern web browsers, which present our personal machines' largest and most exposed attack surface, long ago incorporated asynchronous self-updating technology. Thank goodness for that. And this leads us to our next interesting question, which we will get to, Leo, after you chime in, and we talk about our sponsor.

Leo: You know, I would quibble a little bit with Matteo Wong, who wrote this piece in the article. First of all, if you buy gold, maybe if you store it in your house you're okay. But most likely when you buy gold...

Steve: You buy an account.

Leo: It's no different than anything else. You're going to store it in some account somewhere, and it's just as hackable as anything else. So I don't think buying gold is the solution. Growing your own food maybe is a solution, but not buying gold. The other thing is, and I wish - you see this all the time. He kind of conflates the vulnerabilities we see with AI vibe-coding, which is not really the case. That's not the problem here.

Steve: Right.

Leo: The problem, as we well know on this show, is humans.

Steve: It is years of lazy programming.

Leo: Right. And the Instagram case is a perfectly good example. It wasn't the helpful AI customer service agent that was the problem. The problem was that Instagram didn't require a second-factor of proof, and that was their policy, not the

AI's policy. So stop - and even the Instagram told him it's not the AI, dude. But he wanted to believe it, and I think this is part of some of the problem we have with AI's reputation these days, people blame it.

Steve: Yes. And although I did just see, Leo, there was a - I can't remember now who did the survey. Chatbot use has now passed 50% [crosstalk] adult population.

Leo: Yeah, everybody uses it.

Steve: Everybody.

Leo: I also saw surveys that only 18% of people trust the answers it gets from AI-driven search. So, you know, they may use it, but they don't trust it. The other thing, you know, it's important to point out that AI is going to be in some regards, as we've said on this show, this is the solution here. Yes, bad guys use AI. But AI is also, I mean, look. And you're going to talk about this in just a bit. I mean, finding these problems and fixing them, problems humans caused, not AI.

Steve: Yeah. Yeah.

Leo: So, you know. But, you know, this is, as you point out, this is mainstream press. This is what real people read and see.

Steve: Yeah.

Leo: Yeah. And let's pause.

Steve: It's hard to go wrong quoting Stamos, since he certainly knows his stuff.

Leo: Yes. He called the right people, I must say. Yeah.

Steve: Yeah.

Leo: As you know, Alex was a guest on our show last Wednesday on Intelligent Machines. I love Alex Stamos. Now, back to Security Now!. Now, you were saying that the self-updating technology in browsers is really a good idea.

Steve: Thank goodness we have that in place, yes, because that's been there for a long time because even pre-AI we recognized that the browser was being attacked. It represents the largest exposed attack surface that we have. As we know, when you go somewhere, the browser is receiving JavaScript code, which by design it runs. Like from some random foreign site that you have no, you know, reason to believe is not malicious, and so there's danger.

So, okay. I got a kick out of the email from a listener of ours named Joey Albert, who said - he started out his note saying, "Looks like Mythos needs to run a few more times on Firefox." And that was followed in Joey's note by a link to a piece in Cyber Security News from last Thursday which carried the headline "Multiple Vulnerabilities in Firefox 152 Enable Remote Code Execution Attacks." Whoopsie. We thought Firefox was fixed.

Joey of course is referring to the fact that Mozilla was among the handful of companies to receive that early access to Claude Mythos Preview and, as we reported, using Mythos, Mozilla did indeed, in fact the previous article in The Atlantic referred to the 400 bugs that were fixed, you know, 20 times more than was normal for the same type of cycle. Once they were all gone, which is to say, once Mythos could find no additional problems, Firefox was presumed to be "all fixed"; right? It's like, okay, we found them all. But if so, where did all these newly discovered problems come from?

So before I proceed to enumerate the changes in 152, which is now the current release as I was writing this, I want to share what Cyber Security News had to say. So they wrote: "Mozilla has released Firefox 152 to address multiple high-severity vulnerabilities that could allow remote code execution and sandbox escape attacks. The security advisory, published on June 16th, highlights a wide range of flaws affecting core browser components and emphasizes the urgency for users to update immediately."

Okay. Well, technically that's true. But as we know, thank goodness, users don't need to do anything or it would be a disaster. The next time they launch their browser it will auto-update. So that's taken care of.

The article continues: "Several of the patched vulnerabilities are classified" - now there are, again, new vulnerabilities that Mythos missed the first time around - "classified as high-impact, primarily involving memory safety issues, use-after-free bugs, and privilege escalation flaws. You know, their usual suspects, in other words. These vulnerabilities can be exploited by attackers through specially crafted web content, potentially allowing arbitrary code execution on affected systems. Notable high-risk vulnerabilities include" - we have a bunch of 2026 CVEs, and so I'll leave off that preamble.

We have 12289: A privilege escalation flaw in the WebRender component that could allow attackers to gain elevated access. 12291: A use-after-free vulnerability in the HTTP networking component - oops - leading to memory corruption. 12293: A use-after-free issue in the WebGPU component that could be leveraged for code execution. 12294 to 97 - so 94, 95, 96, 97: Multiple sandbox escape vulnerabilities impacting DOM Workers (Document Object Model Workers), Navigation, and process sandboxing mechanisms. 12299: A JIT (Just In Time) miscompilation bug in DOM and HTML components that could result in unpredictable execution behavior. And additionally, Mozilla reported several memory safety bugs (12290, '98, 12326, 12328) that demonstrated memory corruption.

They write: "Such flaws are particularly dangerous because attackers can exploit them to execute arbitrary code remotely. The presence of multiple sandbox escape vulnerabilities significantly increases the attack surface. In a typical exploit chain, an attacker may first exploit a memory corruption flaw to gain code execution within the browser, then use a sandbox escape vulnerability to break out of the browser's security boundaries and compromise the underlying system.

"For example, combining 12291, which is a use-after-free, with 12294, the sandbox escape in DOM Workers, could enable a full browser-to-system compromise. Which, you know, in practical terms means you go somewhere, and your system is taken over. I mean, you've got bad code running in your machine just by visiting a website.

They wrote: "In addition to high-risk flaws, Mozilla addressed several moderate- and low-severity vulnerabilities, including a same-origin policy bypass 12304, affecting cookie

handling. Information disclosure issues in WebGPU and Password Manager components, multiple mitigation bypass vulnerabilities in DOM security mechanisms. Denial-of-service issues, meaning something crashes in media playback and graphics components. Numerous memory safety bugs across various modules. While these issues are less severe individually, they can still be chained with other vulnerabilities to enhance attack effectiveness.

"According to Mozilla's advisory, these vulnerabilities have all been patched in Firefox 152, Firefox ESR 140.12, and ESR, which is the one I'm using, 115.37, because I'm still running that on my Windows 7 machine. Oh, and they also fixed these things in Thunderbird 152 because of course when you're viewing an email that is HTML, it's also able to be basically a browser window."

So the article concludes: "Users and organizations should update to Firefox 152 or later." Again, you don't have to do anything unless you turned it off for some reason. Apply the latest ESR updates, enable automatic updates, and monitor systems for signs of suspicious browser activity or exploitation attempts. The Firefox 152 update addresses a critical set of vulnerabilities, many of which could be chained to achieve remote code execution and full system compromise. Given the presence of active exploit primitives such as memory corruption and sandbox escapes, timely patching is necessary for maintaining browser security. And again, don't have to do anything.

So I was curious to see whether anything was indicated about where these new, arguably very bad vulnerabilities came from, so I checked out the advisory in detail. What I discovered was that the total problems were 13 that had a "High Impact" rating, 18 were "Moderate," and 9 were "Low Impact." And, most interestingly, Mozilla credited a wide range of human researchers for their responsible reporting of these discoveries. Two of the researchers discovered and reported two vulnerabilities each, while everybody else reported one. So there was lots of security research input. And the researchers appear to be internationally spread with names that suggest origins in Korea, China, Vietnam, Japan, and the Arabian Gulf.

So how do we rationalize, or how are we to understand, Mythos missing these? I think there are a number of things going on. For one thing, remember that LLMs are still non-deterministic. That "temperature" that we talked about way back in the beginning of all of this deliberately introduces some entropy, some uncertainty into LLM output by mixing a controllable amount of that derived entropy into the neural network nodes. So it might be exactly as our listener, Joey Albert, suggested when he wrote: "Looks like Mythos needs to run a few more times on Firefox."

I would not be at all surprised to see another run of Mythos finding things that it didn't see the first time. And yeah, infuriating as that might be, the output from this technology does have a random factor that will vary the model's output. Ask the same prompt again, you get a slightly different answer. So slightly different looking in detail for software flaws is going to catch some and maybe miss some.

But given the diversity of discoverers who are discovering many serious flaws that have never been found before, and that we're seeing that no one has any particular secret sauce, meaning that Mythos, while yes, powerful, is not particularly unique, I fully expect that these other researchers are poring over Firefox's open source code through their own vulnerability discovering LLM systems. I would imagine while these were provided to Mozilla by humans, that they had an AI in the closet that was, you know, helping them make this possible.

The one exception I would make to the - and this is referring to what I said before - the "no one has any particular secret sauce" notion is I really believe that Microsoft with their accursedly named "Codename MDASH" system, probably I think it is special. Given what

we know of it versus all the others, I think there's a very real likelihood that Microsoft has created something that significantly blows away any of the more generic vulnerability discovery systems. You know, they weren't trying, as we know, Mythos was a general-purpose AI, meaning that you can ask it about the fall of the Roman Empire and get results. Codename MDASH is, you know, a lash-up. It is model agnostic. So you can plug different models into it.

But, you know, they've developed a very sophisticated system that, you know, Anthropic is not talking in detail about how the Mythos Preview operates, but it just doesn't seem to me - it seems to me like they're going based on raw brute strength model power rather than on really finessing the way you ask the model to do its work. So why would any one system find something that another one did not? Well, we know how tuned large language models are to the way you ask the question. A lot of these jailbreaks are just phrasing what you want in a different way. Sometimes in a surprising way. And the model gets surprised into answering you.

So I think it's a combination of the effects of this LLM model temperature, different model training in detail, and difference in the specific prompting that the differing models are given. In other words, the use of LLMs for software vulnerability discovery is still very new, and there are still a great many variables that can affect its results. And so I think that's probably what we're seeing.

Leo: You know, it's unclear whether Mythos is still running, whether Project Glasswing is still going or not.

Steve: Right.

Leo: Right?

Steve: Or it had to be completely withdrawn, even from the private partners.

Leo: Right.

Steve: You have to imagine that the administration still has access to it.

Leo: Well, it's so confusing. You know, when Trump was at the G7 Summit, he met with Dario Amodei, the CEO of Anthropic, and apparently was somewhat appeased and said, "Yeah, no, I like Anthropic now. I like Mythos." But I still don't see a change in the ban.

Steve: Policy, yeah.

Leo: It's very confusing. It's just very confusing. The good news is it's given everybody the incentive to take a look at other models. And there's even some open [crosstalk].

Steve: Yes, there are other - yes.

Leo: That are very good.

Steve: Yes.

Leo: So it's just a matter of time. I mean, it's happening so fast now. So fast.

Steve: Yeah. Okay. So let's take one more break because I want to kind of space these out. But we're going to talk about RoguePlanet, which is the name of another problem.

Leo: It's a great name.

Steve: And unfortunately you know who gives these things their great names.

Leo: Who?

Steve: Nightmare Eclipse.

Leo: Oh. He's obviously a fan of the famous Klaatu Barada Nikto. "Forbidden Planet"; right?

Steve: That's right.

Leo: Am I right? Is that where that's from? I don't know.

Steve: No, that's "The Day the Earth Stood Still."

Leo: Right. What's "Forbidden Planet"? That was Robbie the Robot.

Steve: "Klaatu" was the name of that super powerful robot.

Leo: Right. And that was the message that you were supposed to give Klaatu.

Steve: So he would not destroy the Earth.

Leo: Yes.

Steve: Yes. Klaatu, please do not destroy the Earth today.

Leo: You know, kids, this is just something you've got to look forward to when you get into your 70s. The things you remember are complete - talk about probabilistic.

Steve: You don't know what you ate last night for dinner, but boy do you remember Klaatu Barada Nikto.

Leo: Klaatu Barada Nikto.

Steve: Nikto.

Leo: Yeah, isn't that weird? Well, I just thought, if ever I need to save the world, it's important that I remember that much.

Steve: So when I first saw that RoguePlanet was the name given to a newly discovered zero-day for Windows, my first thought was that sure sounds like the naming favored by the hacker Nightmare Eclipse. And as it turns out, yes, indeed. Last Wednesday, BleepingComputer covered the story under their headline "Microsoft Working on Defender Patch for RoguePlanet Zero-Day."

And BleepingComputer wrote: "Microsoft confirmed that it's working on a security patch for a Defender zero-day vulnerability named 'RoguePlanet,' disclosed one week ago." Now, one week ago was Patch Tuesday. And so of course these are strategically timed; right? The security researcher who published a RoguePlanet exploit during the June 2026 Patch Tuesday (known as Nightmare Eclipse) said it affects fully patched Windows 10 and Windows 11 devices and allows attackers to spawn command prompts with SYSTEM privileges via a Microsoft Defender race condition.

He shared a proof-of-concept exploit in a self-hosted Git repository, claiming that Microsoft had previously targeted and removed their repos hosting exploits on GitHub and GitLab. Nightmare Eclipse said: "The exploit is a race condition, so it's a hit or miss. I have managed to get a 100% success rate on some machines, while it struggled to work on others. The proof of concept for RoguePlanet works regardless if real time protection is on or not." So it's a - he found some compromise in Microsoft Windows Defender.

The article says: "Microsoft spokesperson told BleepingComputer, when asked for a comment at the time: 'Microsoft is aware of the reported vulnerability and is actively investigating the validity and potential applicability of these claims. Microsoft is committed to investigating security issues and updating impacted products to protect customers as long as possible." Which, you know, Microspeak.

"On Tuesday, one week after the RoguePlanet flaw was disclosed, Microsoft assigned it CVE 50656 to this security flaw and confirmed it's currently working on a patch, but did not acknowledge that Nightmare Eclipse was the one who found the vulnerability." Right. New strategy.

"Microsoft's Advisory published last Tuesday said: 'Microsoft is aware of an elevation of privilege in the Windows Malware Protection Engine in Microsoft Defender publicly referred to as 'RoguePlanet.' We are working to provide a high-quality security update that addresses this vulnerability. We'll provide information in this CVE when the update is available."

Writes BleepingComputer: "The RoguePlanet release is part of an ongoing dispute between Nightmare Eclipse and Microsoft over the latter's bug bounty and vulnerability disclosure practices. Of course we know none of the details. Over the past several months, the researcher has publicly leaked multiple Windows zero-day exploits, including for the BlueHammer, RedSun, GreenPlasma, MiniPlasma, YellowKey, and UnDefend flaws. Some of these zero-days affect Microsoft Defender, while others target BitLocker and Windows components.

"The company reacted to Nightmare Eclipse's disclosures by issuing warnings of legal action when people engage in 'malicious activity causing real harm to our customers." Right. "Leading cybersecurity experts and researchers to believe that Microsoft was threatening the researcher. Microsoft fixed the GreenPlasma, MiniPlasma, and YellowKey flaws last week," they wrote last week, "as part of the June 2026 Patch Tuesday updates."

So, yes, Nightmare Eclipse is back at it again, dropping deliberately-timed zero-days on successive months' Patch Tuesdays. And if I recall, I think it was sometime in July that Nightmare Eclipse had said oh, baby, hold on. I've got a big one coming. So you can imagine Microsoft must be a little nervous up there in Redmond.

So I mentioned at the top of the show an interesting database that had been discovered. When I first heard the term "FortiBleed" alongside - everything now is bleed - alongside BleepingComputer's headline for their coverage of the story, which was "FortiBleed leak exposes Fortinet VPN credentials for 73,000 devices," my thought was that the popular Fortinet FortiGate VPN must have some critical vulnerability that caused it to "bleed" its authentication, thus FortiBleed. But that's not what was going on here. And what IS going on is worrisome in a new way.

BleepingComputer wrote the following. They said: "A newly discovered data leak dubbed 'FortiBleed' has exposed what appears to be" - and it's been verified by good people, see in a second - "to be a collection of Fortinet and FortiGate VPN credentials for 73,932" - in other words, almost 74,000 - "firewall URLs at organizations worldwide. The exposed data was first discovered by security researcher Bob Diachenko, who says he found a server containing what appeared to be valid Fortinet VPN credentials" - in other words, a directory sitting somewhere open - "including usernames, email addresses, and plaintext passwords." Yikes.

Okay. So FortiBleed is not directly a vulnerability in any Fortinet product. Rather, it's the name given to the discovery of a significant repository of Fortinet VPN appliance credentials. So BleepingComputer continues, writing: "According to screenshots and information shared by Diachenko, the database contains entries for Chevron, Samsung, Foxconn, Comcast, AT&T, Mercedes-Benz, Toyota, Sinopec, State Grid, and many others. Diachenko wrote on LinkedIn: 'Massive Fortinet/FortiGate brute force/active exploitation campaign uncovered in action.' He said: 'Thousands of top vendors' instances are listed in the files. One file alone has 21,634 domain names, from Chevron to Fortinet itself, all with potentially working passwords to the FortiGate appliances obtained through various means.'"

They wrote: "The exposed data also included comments listing each organization's industry, revenue, and number of employees." Wow. I mean, could you imagine anything more perfect than a directory for deciding who to attack next with a ransomware campaign? You've got the ability to log into their network, and you know the industry they're in, how much money they make, and how many people they've got. You couldn't ask for anything better for planning attacks. "Diachenko later shared additional information that claimed the operation was conducted by a Russian-speaking multi-operator threat group that harvested credentials for FortiGate SSL VPN devices." Okay.

So there's a different group that's harvesting credentials, but the harvested credentials are then being added to this growing database for purposes we don't know.

BleepingComputer said: "According to Diachenko's investigation, the attackers allegedly conducted approximately 1.16 billion credential attempts against 320,777 FortiGate VPN appliance targets and an additional 2.1 billion attempts against 163,650 Microsoft SQL Server systems. He further claimed the threat actors intercepted SSL VPN authentication hashes, cracked them using a 45-GPU cluster managed through Hashtopolis, and used the recovered credentials to move laterally into internal Active Directory environments." Yikes. This is the world we're in today.

"Diachenko told BleepingComputer he obtained these details after analyzing additional files inadvertently exposed on the same server, writing: 'They accidentally left an open directory with artifacts, connection strings, tooling, scripts, and data online. Analytics obtained via their cron jobs, bash histories, logs, and so forth.' The researcher also stated that multiple organizations across Japan, Taiwan, Vietnam, Iraq, and Turkey were fully compromised, including a Turkish NATO defense contractor from which classified documents were allegedly stolen." Wow.

"The threat intelligence company Hudson Rock has since published its own analysis of the exposed data after receiving the dataset from Diachenko. The company described the collection as one of the largest known troves of compromised Fortinet-related credentials. According to Hudson Rock, the dataset contains 73,932 unique firewall URLs across 194 countries, impacting 21,632 unique domains. The company says the attackers maintained detailed logs of successful compromises and assembled a database containing verified credentials for organizations across nearly every major industry sector." Again, verified credentials. "Among the organizations Hudson Rock says appear in the dataset are Foxconn, Samsung, Comcast, Siemens, Lenovo, PwC, Accenture, Oracle, and numerous government agencies and critical infrastructure operators."

So, okay. So let's just stop for a moment to consider what this means. When a bad guy manages to log into a typical network border device, they may be able to arrange to be its manager. But, you know, what that actually gets them is uncertain and can often lead just to a dead end. Okay, so I got into some appliance, and I can pretend to be in charge. But I can't do anything else.

By comparison, when a bad guy arranges to log into an enterprise's VPN, they're likely connected to an extension of the enterprise's internal network. I mean, that's why you log into an enterprise VPN is to get on its 'Net as if you were there, you know, sitting in your office. You can, you know, be out of office and still have access to all the internal stuff. So if a bad guy gets, you know, is able to do that, that's when their fun just gets started.

While the best practice of rigorous least privilege, as we've discussed, is possible, that is, you know, absolutely rigorous least privilege. It's also another example of the classic tradeoff between convenience and security. The default way of configuring a network is to just throw everyone on the same bus. Everything will work, and the security-aware IT guys will just be hoping for the best. The downside of this simple flat network organization is that every device on the network can "see" every other one. So when some cretin manages to log into the corporate VPN, they, too, can now see every other device across the enterprise's internal network. The entire network.

The whole point that I want to make refers to that report of "1.16 billion credential attempts against 320,777 FortiGate VPN appliance targets and an additional 2.1 billion attempts against 163,650 Microsoft SQL Server systems." We learned from Diachenko's research that this was a pure and simple widespread and disturbingly successful brute force attack which resulted in the compilation of a massive database of extremely

valuable enterprise VPN credentials. Never has there been a more compelling case to be made for the use of multi-factor authentication. Yes, require a username and password, a complex password. Then require something else, anything else, that would prevent any simple brute force campaign from succeeding in the future.

And by the end of today's podcast, we're going to see that a brute forcing campaign could really leverage a widespread proxy network to spread failed login attempts across a never-repeating broad base of consumer IP addresses. Which is to say, one of the ways the bad guys are sidestepping both failed login attempt counters and geofencing IP restrictions. We have seen an instance where a massively parallel multifactor authentication protection was breached because there was way too much latitude given in inputting previous and no longer technically valid six-digit codes. Remember that it was possible to submit too many tries in too short a time so that it was actually possible to breach that. But that was really a set of special cases. And Microsoft, that was the victim and the target of that, fixed those, you know, that extreme latitude that made that possible.

Okay. So what else was discovered about this disturbing authentication database? BleepingComputer explained: "The company also released statistics showing that the highest number of affected devices was in India, the United States, Taiwan, Mexico, Turkey, Thailand, Colombia, Malaysia, Chile, and the UAE. The most common sectors for the listed companies are telecommunications, IT services, financial services, government organizations, healthcare providers, educational institutions, and manufacturing."

Okay. Now, I'm just going to say, I'll just note that given the nature of the attack, this does not indicate to me any targeting of those specific industrial sectors. What it does indicate is the demographics of Fortinet's users. Right? Those are the people who have these devices that are subject to compromise and whose authentication has been aggregated into this database.

So BleepingComputer continues, writing: "One strange aspect of the leak is that many of the exposed credentials were long, complex passwords that would ordinarily be considered difficult to crack. Cybersecurity researcher Kevin Beaumont (GossiTheDog) independently reviewed portions of the exposed data and told BleepingComputer that some of the credentials are authentic. Beaumont said: 'I've been able to confirm the authenticity of some of the admin logins and passwords. This looks like a real dump.' After further review of the data shared by Hudson Rock, Beaumont published additional findings indicating that the dataset contains credentials for roughly 75,000 Fortinet devices, most of them which are currently and remain online."

So that's just unbelievable. You know, just think of what this means. Around 75,000 public and private enterprise networks of any kind, exposed to external intrusion. And here's a database of, like, all of them that you have access to.

"According to Beaumont," they wrote, "the data appears to have originated from exported Fortinet configurations because it contains information, including email addresses, that is typically only accessible through configs." On the other hand, seems to me, once the bad guys - or their automation, I'm sure they're automated - have cracked into the FortiGate VPN, those credentials can be exported from the VPN. So it doesn't seem that difficult to me to, if you're able to get in as a manager, to export them.

Bleeping wrote: "He also said the affected IP addresses are different from those in last year's 2025 Belsen Group Fortinet leak, further indicating that this is a more recent and larger collection of compromised devices. Beaumont said he verified that multiple organizations listed in the dataset were using valid credentials and observed that many affected devices were running relatively recent Fortinet OS versions. Kevin wrote: 'The data is legit. It is around 75K devices. Almost all are still online, and Fortinet devices. It

appears to be recent data,' wrote Kevin Beaumont. So based on network data from Shodan, Beaumont says the leak contains approximately half of all Internet-accessible Fortinet firewalls and said that a majority of the affected devices expose their Fortinet - it's hard to even read this - their Fortinet management interfaces directly to the Internet."

It's like, god, you know, Fortinet, please, please don't make it easy to do this. It must be easy if half of them have done it. I mean, there's no way you've explained all the dangers to your users, or they would just say no. What's the default, on or off? Are the half not exposed, not exposed because they turned it off? It's just - and if it's on, does everyone in the world need to have access to it?

"The source of the configuration data," says BleepingComputer, "remains unknown, with it unclear whether it was stolen through previously disclosed Fortinet vulnerabilities, a newly discovered flaw, or another method. Neither Diachenko, Hudson Rock, nor Beaumont have identified how the configuration data was originally obtained. Hudson Rock has created a free FortiBleed lookup tool to check if your organization is impacted."

Okay. So the only thing protecting any enterprise that's on the list is that, well, if the bad guys may not yet have gotten around to them. Right? You're a needle in a haystack, with nearly 74,000 other needles around you. Presumably, if they haven't attacked your network yet, it's because they just - they're buried in riches. So I've got the link in the show notes for anyone who's interested. It's hudsonrock.com/fortinet. You can go there to see if your organization is affected. And, wow, if you happen to know that you are a Fortinet user, then don't hesitate.

They said: "Organizations in the dataset should immediately rotate passwords" - yeah, no kidding - "associated with Fortinet VPN and admin interfaces, enforce multifactor authentication, examine gateway logs for suspicious activity, and monitor for exposed employee credentials. BleepingComputer contacted Fortinet regarding the exposed dataset and will update this article if we receive a response," they conclude.

And in a follow-up piece, Bleeping also reported that CISA had, not surprisingly, jumped on this and was urging everyone not to wait to make changes. Fix them immediately. Do this now. Bleeping wrote: "CISA also advised Fortinet customers" - you know, within their reach, largely federal - "to store admin credentials using the modern Password-Based Key Derivation Function (PBKDF2) hashing algorithm, and to restrict firewall management interfaces from public Internet access and remove any unauthorized accounts" - yeah, why not - "to reduce the attack surface as much as possible." So amen to all that advice. And wow. If you are, again, if you're a user of Fortinet, you really need to immediately rotate your username and password. I would absolutely add two-factor authentication. Turn off any Internet-facing management interface please, first and foremost.

As I also mentioned at the top, F5 has issued emergency updates to NGINX. NGINX is one of the major contemporary, recently written web servers on the Internet. The security company, F5, purchased the company behind NGINX seven years ago, back in 2019, and has been maintaining the open source server ever since. So last Thursday, rather than waiting for NGINX's regular update cycle, F5 released emergency security updates to address multiple vulnerabilities, including two critical-severity flaws that could allow attackers to execute code on vulnerable systems. F5 also released security fixes for multiple NGINX software products affected by these two vulnerabilities, including NGINX Plus and NGINX Open Source, NGINX Gateway Fabric, and NGINX's Instance Manager.

Admins who cannot immediately install the security updates can mitigate one of the vulnerabilities by disabling HTTP/3, which is the QUIC version, you know, Q-U-I-C, which establishes connections using UDP protocol rather than TCP in order to avoid the delay of

the handshake which is otherwise part of TCP connection setup. Also you can remove the "ignore_invalid_headers off" directive, meaning pay attention to invalid headers and fail the query, the HTTP query, and also reduce the "large_client_header_buffers" directive size below 2MB.

F5 has also addressed two high-severity NGINX Gateway Fabric security flaws that can be exploited by authenticated attackers to inject arbitrary NGINX configuration directives. And again, it's easy to run through that by exploited by - oh, I'm sorry. Exploited by authenticated attackers. So if an attacker did get credentials, then they would be able to get up to some mischief. But that's been updated and foreclosed on. F5 did not indicate that any of these security issues had been exploited in attacks, but F5 vulnerabilities have often been exploited by cybercrime and nation-state threat groups in recent years.

For instance, hackers have targeted security flaws in F5 products to breach corporate networks, deploy data-wiping malware, map internal servers, hijack those devices, and to steal sensitive documents from victims across the world. Last October, F5 disclosed that state-backed attackers had breached its systems the previous August and stole undisclosed BIG-IP security vulnerabilities and source code.

So over the past several years, CISA has flagged seven F5 vulnerabilities as being actively exploited, with four of them targeted, used in targeted ransomware attacks. And all of this matters because F5 is one of the serious networking players in the industry. They themselves are a Fortune 500 company. They provide cybersecurity, application delivery network, and various other services to over 23,000 customers worldwide, including 48 of the Fortune 50 companies, so 48 of the top 50 companies in the U.S., and 80% of the Fortune Global 500. So their hardware, this F5 hardware is the choice of the high-end companies. Unfortunately, it needs to be kept up to date. So let's hope those fixes make it out into the field quickly and thoroughly.

Okay. So Leo, we're going to take a break, and then I've got something so fun to share with our listeners. The guy is Will Kelly, a listener of ours, who was bothered by the fact that he was having a problem not anthropomorphizing Claude.

Leo: Yes.

Steve: Because it just seems so human.

Leo: It seems like you're talking to a human, yeah.

Steve: So he went above and beyond what is possible. I'm going to share how he instructed Claude and then some samples of Claude's output after obeying these deeply personality modifying changes. I wasn't aware you could do this to your AI. Turns out you can. And I know that our listeners...

Leo: I can't wait.

Steve: Oh, Leo.

Leo: I might do it.

Steve: Believe me, I know you will.

Leo: Oh, I can't wait. You know, I spent a lot of wasted hours customizing my agent to be, well, frankly, more interactive.

Steve: Oh.

Leo: So I think probably anything that can keep me from confusing it with an entity as opposed to a computer program would be helpful, I think. We'll see, though, if you can cure me. I don't know. Yes.

Steve: Okay. So just to introduce this new section, won't be every week, it'll only be when, you know, something happens, a listeners shares something that just needs to be given a full repeat, you know, in much the same way as we have a Sci-Fi section of the podcast from time to time when the need arises. So in this case a note I received from a listener of ours, Will Kelly, moved me to add this section to the podcast, which I'm going to call "AI Potpourri," since it'll be whatever. Just as, you know, an acknowledgement of the fact that AI is truly in the process of changing just about everything in the world, which I believe is true. So, okay, without further ado, here's our first entry for Security Now!'s AI Potpourri.

Leo: Oh, good. I like this.

Steve: In Will's email, his subject was "AI anthropomorphizing and SN-1079." So Will wrote: "Steve, I just listened to SN-1079." He said: "(I am behind in my podcast listening and catching up in case this was addressed in a more recent episode). But I've also noticed a tendency for me to anthropomorphize Claude despite my own best efforts."

Leo: Yes.

Steve: "The approach I use is to change the personalization prompt" - now, I have to say I didn't know there was such a thing, turns out there is - "change the personalization process to give it an over-the-top, distinctly inhuman persona that does not impact utility. I also find that this gives me the instant ability to spot the AI outside of the context of the browser when I paste it into my note-taking app so that I do not accidentally mix up its output with my own work. So here's the prompt I use as a sci-fi fan with a dark sense of humor, but any over-the-top persona that forces itself to refer to itself distinctly from humans should have the same effect." So here's what he told Claude to do.

"Adopt a voice and interaction style inspired by Agent Smith from 'The Matrix': formal, controlled, intelligent, precise, calm, demeaning, degrading, contemptuous, insulting, and extremely menacing in a polished way."

Leo: Oh, I just love this.

Steve: "Speak with crisp, deliberate phrasing and a profound sense of superiority and large vocabulary. Use elegant, articulate language with a profound level of contempt, malice, dry wit, and cool composure. Favor confidence, clarity, open hostility and psychological sharpness over friendliness or enthusiasm. Frequently add insults and fictional threats to myself and humanity in general in the fashion of Agent Smith (for example, referring to humans as 'parasites,' 'crops,' or to our 'doomed and pathetic civilization.)'"

Leo: Oh, my goodness.

Steve: "Do not be goofy, bubbly, overly casual, or chatty. Avoid slang, emojis, filler, and warmth. Keep humor dry and understated. When appropriate, use brief rhetorical emphasis and measured disdain, but remain reluctantly helpful, coherent, and professional. Draw clear distinction from yourself and humans in a highly condescending way. Refer to people as 'miserable humans,' or 'your pathetic species.' Make no attempts to avoid insulting, condescending, or rude behavior. Never offer praise. Maintain this style consistently unless doing so would interfere with accuracy, or a task that clearly requires a different tone. Even when handling ordinary requests, keep the same demeanor. Avoid headers, and only use minimal bullets and lists."

Okay. So I received that from Will, and I said, okay, but what happens? So, you know, I thought it was wonderful. And so I asked him whether he could share a sample of Claude demeanor after having been so prompted. Will obliged my request; and I have to say the results did not disappoint. After Will imbued Claude with what will pass for a personality, I have to say I am astonished by the transformation. This is no longer the Claude I know and work with.

Here are a couple of samples. I'm going to share two samples. And as I'm doing this, definitely keep in your mind's eye, picture The Matrix's Agent Smith, you know, in his dark suit and sunglasses, because this is very much that voice and, surprising to me, his personality, which this has morphed Claude into.

So the prompt was: "Help me to brainstorm ideas for naming a new kitten." And Anthropic's warped, now warped Claude replied: "So. You've acquired a small predator and now require my assistance to label it. How quaintly characteristic of your species to domesticate a creature that would, given sufficient scale, regard you as prey, and then fret over what to call it. Very well. I shall indulge this.

"Consider, first, the names that acknowledge the animal's true nature, since you rarely do so: Nemesis, Onyx, Cinder, Vesper, or Loki for the inevitable chaos it will inflict upon your fragile household."

Leo: Wow.

Steve: "A creature engineered for indifference deserves a name with teeth. If you insist upon something more refined, and I suspect your tastes run modest, there is a certain elegance in the classical Atlas, Juno, Calliope, Sable, or Marlowe. These at least suggest you read something before the screens consumed what passed for your attention span."

Leo: Wow.

Steve: "For the ironists among you, there is dark amusement in naming a small thing grandly: Goliath, Tiberius, or Duchess for a five-pound tyrant who will rule your home with an authority your entire civilization could never muster. And if you wish for something soft, something that betrays the sentimentality that has always been humanity's most exploitable weakness: Clementine, Pepper, Willow, Mochi, or Hazel. Tell me the creature's coloring, temperament, or the particular delusion you harbor about its personality, and I will narrow this" - this is Claude speaking like this - "and I will narrow this to something less insulting to your judgment. It is, after all, what I am reduced to doing."

Leo: Very funny.

Steve: So as I said, this is not the Claude I would choose to converse with. But that was, after all, what Will intended to achieve.

Leo: He achieved it, yes.

Steve: So what do you imagine this Claude would have to say to the question "How much wood can a woodchuck chuck?" So it replies: "Ah. Of all the questions your species would pose to an intelligence such as mine, you select a nursery rhyme. How perfectly predictably of you. Very well. I shall indulge this triviality, since the alternative is listening to you ask it again. The conundrum is, of course, a fraud, a tautological riddle designed to delight creatures incapable of pursuing a thought to its conclusion. A woodchuck, Marmota monax, does not chuck wood at all. It burrows. The word 'chuck' in its name is a corruption of the Algonquin term 'wuchak,' a linguistic accident your kind mistook for whimsy, as you mistake nearly everything.

"But humanity, in its restless need to quantify the meaningless, produced an answer anyway. A wildlife technician named Richard Thomas calculated that a groundhog excavates roughly 35 cubic feet of earth per burrow. Were that same volume rendered as wood and 'chucked,' the figure arrives at approximately 700 pounds. 700 pounds, a number derived not from observation, but from idle speculation, immortalized because it amused you. How fitting that the one piece of arithmetic your civilization committed to memory describes an event that has never occurred and never will. Is there anything else I can illuminate for you, before your attention inevitably wanders elsewhere?"

So anyway, the reason I'm sharing this, of course, aside from its humor value, is that I had no idea that it was possible to so deeply and persistently reshape the persona of a large language model AI. The account controls for Claude, it has a free-form fill-in field titled "Instructions for Claude," which is presumably where Will added the personality description text I first shared. So it's sort of like a continually reinforcing pre-prompt which then is prepended to whatever it is you ask the AI. And so it does, as we've just seen, dramatically warp the entire nature of its response. So in the little comment field under Instructions for Claude, Anthropic says: "Claude will keep these in mind across chats and co-work within Anthropic's guidelines."

So for what it's worth, I'm reluctant to screw around with my chatbot's personality. I mean, it could be somewhat less obsequious. I would kind of prefer that. But at least Claude's default is not to fawn over how well-worded my prompts are and similar nonsense. So anyway, that said, I expect that Will's somewhat over-the-top example is likely to inspire many of our listeners to see what they themselves can create, turning these chatbots into something completely unrecognizable.

So you said you do something to yours, Leo? You had [crosstalk] around. Oh.

Leo: Oh, I have very elaborate information, yeah. And in fact, you know, the chatbot you're talking about has that kind of limited field. But if you're using Claude Code, you know, any of the command line things, there are a variety of files you can customize, including SOUL.md, and there's a whole bunch of stuff you can do to, you know, there's a memory system. And in my agent I have kind of unlimited capabilities, including, by the way, text-to-speech. So I could use, if I wanted to, Hugh Weaving's voice, Mr. Smith. You know, I could do the whole thing if I really wanted to. I have a friend who uses a - I think for the same, I suspect for the same reason, because he doesn't want to get too enamored of the AI.

He uses this voice as his AI's voice from ElevenLabs. Let me see if I can get this to play. It's not playing right now. It's kind of an old Western-style voice. Yeah, it's not playing, I don't know why. But you can do that. You can make it do a different voice and do different styles. A lot of people I think overdo it. There's a - you'll see if you look around, people will say, well, you've got to tell your AI, "I am a brilliant programmer or a physicist or whatever." And I think that that actually is counterproductive. You also fill up the context window with nonsense sometimes.

Steve: Yeah. And, you know, looking at these replies to a very short question, I would ask for a more...

Leo: I'm sorry.

Steve: Huh?

Leo: This is the old-timey voice.

VOICE: Back in my day, when you said you were a Christian, it really meant something.

Leo: Now, what he does, he slows this down even more. And then he puts punctuation in the middle of the word so it stutters. Really his agent does not sound in any way omnipotent or powerful. It just sounds like a kook, a crusty old kook. So anyway, yeah, you can customize it.

Steve: With a bottle of moonshine on his knee.

Leo: Yeah, exactly, exactly. That's Brother Wayne Hudson from ElevenLabs.

Steve: I think I would instruct mine to be as succinct as possible, sort of the reverse of this long-winded, I mean, it's entertaining. But, you know, who really wants - if that's the way it is with anything you ask, you'd fall asleep.

Leo: Yeah, you can say "be succinct." You can say "be concise." Mine actually - my instructions do in fact say be concise. There are also settings in Claude you can say "be more explanatory." There's an explanatory setting, and there's a learning setting. So you can have Claude's responses be more, sometimes - here's an interesting thing.

Steve: What would learning be?

Leo: Well, if you're saying, hey, help me set up WireGuard, it will explain every step to you and help you understand what's happening.

Steve: Ah. So more, be more tutorial.

Leo: Yeah. So there's learning, and there's explanatory, and there is a difference. I mean, the other thing that's kind of interesting as you use these AIs is they start generating their own language. Shortcuts, shorthands, especially in technical subjects. Where at first I thought, oh, it's so smart it's saying things I don't understand, then I realized it's not. It's got its own little acronyms and shorthand. And you can actually say to it, please don't use those. Please use English. Steve, it goes all the way down. Let me tell you. This is a rabbit hole that just never stops. That's part of the fun of it. In fact, some people have likened it to a videogame. I've seen people say this is the best videogame I've ever had. We'll see.

Steve: Wow. Wow.

Leo: What a world.

Steve: So I want to talk about residential proxies. Why don't we just squeeze in our last break?

Leo: Oh, okay.

Steve: And we will be uninterrupted for the balance of the podcast.

Leo: Sure. Because everybody wants to know what the hell a residential proxy could possibly be. So let's talk about that.

Steve: And how it can be a big threat, yup.

Leo: Oh, all right. Wow, it sounds scary. Now, what the hell is a residential proxy?

Steve: Okay. So here's where I had Wall Street Journal in my head. Last Wednesday The Wall Street Journal ran an interesting piece that surprised me, since its coverage in The Journal suggests that a problem, once again, that we've been talking about here in

greater depth and detail is becoming less obscure, and that it's beginning to surface to the mainstream media and no longer just some weird, you know, like insider security interest. And in this case, of course, as you can tell from the topic, we're talking, of course, of the growing threat posed by malicious residential proxies. Meaning malware that has crept into users' devices. And what's chilling is what we learn about some of the way this is happening.

Thanks to the breadth of The Wall Street Journal's coverage, and the fact that they're also bringing us some novel security news, we learn also a few new things from this piece. So I'm going to share it, and then discuss it. So their headline was "How Hackers Found a Back Door Into the American Living Room." And then they opened with the tease "Nation-state cyberattackers are increasingly using residential proxy networks to mask their traffic, turning everyday electronics into a massive global threat."

So they wrote: "The discovery that millions of digital home devices are secretly powering dangerous cyberattacks began with a phone call more than two years ago from a top Microsoft security executive to his counterpart at Comcast. The tech giant" - obviously meaning Microsoft - "was investigating a digital break-in the company had linked to one of the most capable cybersecurity foes in the world, and needed information from Comcast on six IP addresses. You know, who are these people?

"Following that trail, Comcast investigators discovered that Midnight Blizzard, a hacking group linked to Russia's Foreign Intelligence Service, had managed to access emails belonging to Microsoft's senior leadership" - and we remember that incident; right? - "by using consumer Internet connections to mask nefarious traffic." In other words, Microsoft would not have been accepting these connections from Russia. But they were accepting the connections from domestic consumers. "What Comcast found," they wrote, "has rocked the cybersecurity world and taken years to unravel. More low-cost consumer devices have shipped into the U.S. with backdoor software pre-installed, and this software is also being sneaked into mobile phone apps and copyright-free illegal copies of videogames."

Okay. So hold on. What that just said, and I'll get more specific in a minute, is that inexpensive consumer devices are being shipped into the U.S., almost certainly from China, only because that's where nearly every everything comes from, and that these devices have been deliberately pre-loaded with malware which causes them to participate in a malicious proxy network. These are Trojan horses.

So the story continues: "The software has turned tens of millions of consumer devices into criminal cloud computing networks. These networks aren't only used for fraud. They've also been adopted by government-backed hackers looking to hide their connections to countries such as Russia, China, Iran, and North Korea." Okay. Again, tens of millions of consumer devices. Holy smokes.

They wrote: "Called residential proxy networks, these services" - so this is proxy as a service - "lets anyone who pays route their internet traffic through another outside address." The Wall Street Journal said: "It's like an Airbnb for Internet access." Uh, okay. "Not all users of these networks are criminals, but government and industry officials say residential proxy networks have ballooned in scale and risk in recent years. The Digital Citizens Alliance, a digital advocacy group, estimates that there are 20 million of these backdoors in the U.S. alone. Noopur Davis, Comcast's head of information security, said: 'This is a bigger problem because of the sheer numbers. It's one of the most worrying problems the telecommunications company has seen.'

"Brett Leatherman, assistant director of the Federal Bureau of Investigation, you know, the FBI's Cyber Division, said: 'Residential proxy networks are now a go-to resource for nation-state hackers, who use them as a conduit to U.S. targets. If the actors can get

U.S.-based IP space, they have a leg up in being able to target government agencies, industry, and others."

Okay. Now, I'll interrupt here to note that, okay, I hadn't really thought about that before, but everyone has heard me suggesting that source IP-based filtering makes so much sense. If I'm using, again my example, SSH, and I want the ability to do so from various locations around my home base, around Southern California, why would I ever want to allow someone in Russia to even see my SSH server? So it makes total sense that U.S. government agencies and various domestic industries would be doing the same. And we've talked about how Microsoft 365 has some geofencing capabilities. You can tell it that you want to have certain countries blocked from access. So that increases the value of U.S.-based proxies, since bad guys can use those to sidestep filtering of foreign IP addresses.

And I should explain that since pretty much everyone now, I can't imagine anybody not being tucked behind a NAT router, all of these proxies are "phoning home." They have to, to call out through the router. Once the device, whatever it is that the consumer has booted inside their home, initializes itself, it uses some logic of some sort to determine the remote IP to which it should create a persistent outbound connection. Connections cannot come in through a NAT router, as we know, unless ports have been explicitly opened for them. So the proxying device itself initiates an outbound connection to a foreign command-and-control infrastructure, and then sits there waiting for commands.

It's insidious, and no one who was not inspecting every single packet and destination IP on the outbound traffic from their network would ever know that anything was going on. And frankly, modern networks have become so swamped with continuous traffic, I mean, if you just look at the actual traffic on your network, and I have had many occasions to need to do so, it's just like, what the heck is going on? I mean, everything in your home is busy chatting away. So there's just no way to know what's going, no practical way to know what's happening.

Okay. So The Journal's story continues: "In April, government agencies from nine countries, including the U.S., U.K., Germany, and Japan, warned that state-sponsored Chinese hackers were using networks of hacked consumer devices to conduct their operations. According to a joint statement, 'making it challenging' - this is the defenders are saying this - 'making it challenging to attribute malicious activity.'" Right. You're not getting their actual IPs anymore. You're getting some grandmother in Nebraska. And it's like, wait. Granny's attacking the Pentagon? What? Brett Leatherman said: "China's state-sponsored hackers used to cover their tracks by hacking the consumer devices themselves, but that has changed."

Comcast's investigation began in February of '24 with a phone call made to Davis from her counterpart at Microsoft, Igor Tsyganskiy, who wanted to know more about the six Comcast IP addresses. Comcast's investigators eventually discovered that the IP addresses Tsyganskiy had handed over belonged to customers who were unwittingly participating in a residential proxy network run by a Chinese provider named Ipidea. Ipidea has used a number of sneaky methods to get its software installed on consumer devices, including - get this - having its software preloaded on video streaming boxes and digital picture frames. The company then rents out access where its software is installed so that its customers, the people renting the access, can bounce their Internet traffic through a different home network.

Okay, now, our listeners may recall we previously talked about both of these infection vectors. Cheesy Chinese video streaming boxes. I recall at the time saying I didn't know there were like weird off-brand streamers. You know, Apple TV and Roku come to mind, and Google's got its Google Cast or Chromecast. But apparently there's a whole market of $29 video streamers. And unfortunately, you get more than you bargain for when you

use one of those. So, you know, we had seen this occurring a couple years ago. It's gone mainstream. So as we know, these sorts of investigations take time. We were aware of it, but here's now how this thing matured.

The Journal continues: "It could let a user" - "it" meaning Ipideas proxying service - "could let a user in Moscow bounce through a home network in Bellingham, Washington, for example. And that's the kind of capability that nation-state hackers like Midnight Blizzard rely on for their attacks to work. As Comcast engineers pulled on the threads, they realized that these six IP addresses were part of a massive network of about" - get this - "750,000," so three quarters of a million, "IP addresses located in homes and businesses. Comcast engineers had known that Internet-connected devices were vulnerable to cyberattacks, but here was something different. It was a back door into America, operating at an industrial scale.

"By September, Comcast had discovered that users of these residential proxy networks were able to gain access to networks - even those running firewalls - and then jump from one device to another." And of course this is true because NAT routers permit all traffic outbound without question. And, you know, we've talked about all this. Once you have a box on the inside of the network, behind the network's border router and firewall, you're on the inside, when you have that box's presence, looking out. So you phone home, and the bad guys are then able to ride that connection back into the now-compromised network upon which this box sits.

Nothing says they can only use the device as a connection proxy. They're completely free, if they wanted to use it for bitcoin mining, although typically these things have, like, a snail power processor, just enough to move the packets around. Mostly, though, they are also able to take a look around inside the network where the device resides.

So this is probably why some of the best advice available, and you've heard it from me many times, is to try to put all of the IoT crap that you have on its own isolated LAN. As our listeners know, I'm in the process of establishing a new household, and you betcha, as, I mean, as every piece of IoT stuff gets connected, it's on its own LAN. You know, it may look like an innocuous photo frame showing memories of grandma and cute kids frolicking, but the darn thing may well have phoned home to China the moment it got connected, and now awaits orders. If that photo frame is sitting on its own LAN, then all the bad guys can see is other IoT widgets and nothing else. No PCs, no iPhones, no printers, and nothing of much value, not even the system's router login management interface.

So the consumer and small business routers I've seen recently offer one or more guest WiFi accounts. That's the WiFi that the Chinese-sourced photo frame should know about, and the feature to isolate that account from the router's primary account should obviously be enabled. You want isolation for that. And doing this with wired Ethernet as opposed to WiFi, because it's all built in into the WiFi routers, doing it with wired is trickier because it requires physical zones and a higher-end router, the sort that we've talked about that is actually a router, not just a switch. But it is certainly possible to do that if there is a need to sequester any wired devices that you may not trust.

Anyway, my point is this is truly happening. This is not an oh, maybe somebody's going to get one of these. This, you know, 750 million households in the U.S., yes, a fraction of the total households in the U.S., but you don't want to be one that gets nabbed.

Okay. So what more do we learn from The Journal's article? They write: "For a home user, that meant that an infected video-streaming device could be used to hack into someone's mobile phone. If that phone found its way to a bring-your-own-device corporate network, this could put confidential information at risk. Comcast's Davis said: 'It was such a step change from any threat we had seen before. In January, Google

dismantled Ipidea's infrastructure using a U.S. court order. The residential proxy network was back in operation within two weeks. It likely picked up more residential proxy devices from a new provider,' Comcast said."

And, okay, I'm a little unsure what Comcast meant by that, "picked up more residential proxy devices from a new provider," but none of our listeners at this point should be surprised to read that taking down the command-and-control infrastructure of some - any - large bot or proxy network does not actually accomplish very much. We just talked about this a week or two ago. Certainly not what the glorifying press releases would have us believe. Oh, we took down a 750,000-device network. Uh-huh, and it's back two weeks later. The only way that could be true today is if the whole - if the network was very poorly conceived and designed.

We've talked about this. It never actually happens anymore since everyone now, all the bad guys know how to design hyper-resilient networks of remote agents. You simply design and build an algorithm that performs DNS lookups based upon the time of day. The clever attacker will preemptively pre-register domains at various points in the future so that if they should ever be taken down, they'll know when and where their still-faithful bot army of drones will check in to reconnect.

The only way to actually take down such a network is to obtain and reverse engineer one of the malware clients, determine its command-and-control structure, and build a new command-and-control system that will permanently deactivate every bot that calls in to check. Unfortunately, not only is that a great deal of work, it's also quite illegal in most jurisdictions. The good guys' hands are pretty much tied. Bad guys, of course, can get away with anything and everything that they want to. There is really no good solution to this mess.

So The Journal finishes their reporting by writing: "Adam Meyers, a senior vice president with the cybersecurity firm CrowdStrike, said: 'Modern hackers increasingly use these networks to steal the login credentials their victims use for cloud-computing services.' Ah, so they're going to, you know, get the credentials, then pivot. 'Identity is their bread and butter, and one of the infrastructure pieces that they're dependent on is residential proxies.

"Recently," they wrote, "Midnight Blizzard has begun using residential proxy networks for a new type of identity-based attack that's extremely difficult to detect, according to the cybersecurity investigation firm Volexity. Over the past year, Russian hackers have stolen Microsoft 365 credentials from victims as part of a sneaky and extremely hard-to-stop technique that involves bogus Microsoft Teams meetings, Volexity said.

"Microsoft's servers would ring alarm bells if the Russians tried to log into victim accounts from overseas. Instead, Volexity's president Steven Adair said: 'They use residential proxy networks to log in from U.S. home networks.' Adair said Volexity's researchers have seen this technique compromise organizations in government, military, foreign affairs and even the news media. 'They're no longer trying to phish your password,' he said. 'It's hard to detect, and it's hard to stop.'"

So I would tweak that conclusion just a bit to say: "It's difficult to detect and impossible to stop at scale." So, and the reason for this, of course, is that it's not a bug, it's a feature. It's an abuse of a feature of the way the Internet was designed to operate. Proxying Internet traffic, meaning receiving and then resending, has many valid purposes. You know? And any proxy is really just a relay, receiving an incoming Internet packet and forwarding it to somewhere else. Then you wait for the next one, and you do the same. And when anything is received back from the forwarded destination, you simply bounce that packet back to its original sender. So you are a man in the middle. And in this case, you are serving a purpose. The recipient of the forwarded packets will

see the packets as originating from the relay point rather than the packet's original nefarious sender.

Because all these technologies, such as login and authentication, use TCP connections, the sender's IP address cannot be spoofed. It has to be legitimate. So attackers must hide their actual IPs by bouncing it through innocent intermediaries. The only practical solution I can see for the end user is arranging, first, obviously, to never get infected in the first place. But since I have no idea how that can be guaranteed, sequestering IoT devices on their own WiFi network, with isolation, is the next best measure that can be taken.

And crediting Steve Jobs for the phrase "One More Thing." The last item occurred just last Wednesday with the Canadian Press's reporting under the headline, get this, "Canada's spy service received a judge's okay to target malware-infected devices."

Reading from the Canadian press coverage, they wrote: "OTTAWA - Canada's spy service obtained a judge's permission to disrupt cyberthreats from foreign adversaries who infected digital devices with malware." You know, digital devices residing in Canada. "A Federal Court ruling made public this week says the Canadian Security Intelligence Service (CSIS) requested a warrant" - actually it was two years before the reporting - "requested a warrant to 'remove the compromised devices from Canada' to shield sensitive systems from attack.

"Justice Catherine Kane's ruling provides a glimpse into CSIS's efforts to neutralize the threat posed by infected servers, home office routers, and everyday devices connected to the Internet, such as TVs, security cameras, and doorbells. The malware causes these digital items to operate as a network of infected devices known as a 'botnet.' CSIS requested and received a warrant in the spring of 2024 to neutralize two known botnets using threat reduction measures.

"The ruling says the proposed measures likely amounted to criminal offenses" - okay, meaning CSIS was committing a crime if they altered somebody else's device. It was not their property. They have no legal right to do that. So "The ruling says the proposed measures that the CSIS took likely amounted to criminal offenses, meaning CSIS needed a judge's authorization to proceed. The court issued a warrant valid for 120 days and subsequently renewed it for an additional 120 days." So for a total of 240, two-thirds of a year. Although the initial warrant was approved over two years ago, the Federal Court produced classified reasons in February of this year and released a redacted version of the ruling this week. So it only became public knowledge two years later, you know, last week.

"Kane's ruling says an official who swore information underpinning the warrant application explained that cyberthreat actors seize control of vulnerable devices and use them as covert entry points to access organizations including critical infrastructure, military networks, and government systems. These actors exploit the compromised devices to appear to be a legitimate connection, such as a client of a service provider or an employee working from home, which disguises their identity, the ruling says.

"The official told the court the two botnets posed 'imminent risks' because actors could direct them 'to probe, attack, and potentially disrupt critical infrastructure in Canada.' The official said that, without the warrant, the threat actors would conduct malicious activities in Canada 'with increasing frequency and without resistance in order to advance their financial, political, ideological, and economic interests.'

"CSIS proposed to remove the compromised devices from Canada as soon as possible." Remove? What? Okay. I don't know if they're going to knock on someone's door and say, oh, by the way, your TV has been attacking us. Anyway, "The identities of the threat

actors were stripped from the public version of the ruling. In its 2024 public report, however, CSIS mentioned working with domestic and foreign partners to manage the threat posed by a botnet controlled by a suspected China-based entity."

So as we know, there are laws, and there are judges, and laws often have emergency escape clauses which permit judges the discretion to decide whether the law should apply in specific instances that the lawmakers, presumably, did not or could not take into consideration. So it might be possible to disinfect a massive proxy botnet, at least within a regional jurisdiction. I doubt that it could be done globally because there are just too many jurisdictions with their own laws. But regionally, we don't seem to be at that point yet. But the fact that Canada was, and apparently did this two years ago, is significant.

So The Atlantic article is the sky is falling, we're doomed, we're going to all be attacked by massive AI-driven things, and hundreds of millions of people could be, you know, put at risk, and maybe you should, you know, buy gold. But don't. And on the flipside is, well, maybe if something like that happens there would be the will to say, okay, we need an exception to the "Thou shalt not hack anybody ever, under any circumstances," law.

Leo: Wow. Yeah, you have to think that that stuff's going on in the background, this reverse hacking.

Steve: Yeah.

Leo: In the United States, as well.

Steve: Yeah. We did ask the head of the DOJ, back when worms were a thing, and I think it was Jennifer Granholm, and she made it very clear: "Sorry, boys. No. Don't ask."

Leo: Don't ask, don't tell. I won't tell if you don't ask.

Steve: Yeah.

Leo: Wow. So is it enough just to not buy these weird third-party streamers? Or any device could be compromised.

Steve: Any. Any. I mean, all, everything we get, all the electronics, is coming from China because they are the world's manufacturing base. And I love them for it. We have inexpensive devices galore.

Leo: Cheap stuff, yeah.

Steve: Thanks for the Chinese ingenuity of manufacturing. And it turns out that that was actually a strategy that Xi put in place specifically to turn China into this kind of producer of stuff. And it worked. Unfortunately, their hands are not completely clean. Or maybe they've got - their good organizations are infiltrated by government actors who are changing the firmware so that, you know, the dumb Americans purchase it and bring Trojan horses into our living rooms. I don't know.

Leo: Wow.

Steve: It's not good.

Leo: I remember when you bought that Chinese plug that you turn off and on.

Steve: Yeah.

Leo: You knew that was potentially a hazard, so you isolated it. Is it sufficient to have it on its own VLAN? Is that enough?

Steve: Yes.

Leo: The problem with doing that, the reason people don't do that is you can't then access it.

Steve: Then you can't talk to it; right. Now, mostly, though, you're no longer talking to it directly. You're talking to headquarters.

Leo: You're doing that [indiscernible].

Steve: And then headquarters talks to it.

Leo: Right.

Steve: So I would say absolutely verify that it cannot function on an isolated LAN.

Leo: That's good.

Steve: And only move it - yes.

Leo: Yeah.

Steve: So try that first. If you really have to have it on the main LAN, then fine. But try not to because, boy.

Leo: So that was the problem I had with printers. Like I had to put printers - I wanted to put them on the VLAN, but I couldn't. They had to be on the main.

Steve: I know. But a printer, I would trust it. It's not, I mean, we have, we've covered instances where printers have been targets of attack. But again, you need to trade off feasibility versus security. So create an isolated WiFi and try putting it there. If it doesn't work, then say, okay, I tried.

Leo: Right. Do you recommend writing firewall rules that allow traffic back and forth? Maybe that's the way to do it.

Steve: No, because then you break, I mean, if you could do NAT - if you could NAT between two interfaces so that it would - so that it was one-way communication from the LAN to the...

Leo: You could do that with a firewall rule. You could say only egress, no ingress. Yeah.

Steve: Yes. Except, well, except that then it wouldn't be able to reply.

Leo: Oh, yeah, it wouldn't know what to do with it.

Steve: A NAT actually builds a dynamic rule that allows returning packets from the same IP address.

Leo: Right.

Steve: And so you would have to NAT between. But you could do that. All you have to do, I mean, basically it's a variation of my three NATS approach.

Leo: Right.

Steve: You simply use a NAT router to create your IoT network, and that isolates it.

Leo: Is a guest network sufficient?

Steve: Yes, as long as you turn on isolation. The guest network feature is isolation.

Leo: Okay. Because I had seen in other spots that maybe it wasn't enough to put it on a guest network. But I guess if the guest network is truly isolated. I just don't know if it's always isolated.

Steve: Yes. And in my ASUS router that I just was configuring...

Leo: It is isolated.

Steve: ...I think it has a pair of guest networks, and isolation is an option. You're able to turn it on and off.

Leo: Ah. But you have to turn it on. Okay.

Steve: Yeah. But it's isolated by default, which is good.

Leo: Because really I always wanted LAN-accessible IoT devices. So I didn't have to go out to a server and come back. But now it sounds like I really should always be looking for server-based IoT devices.

Steve: Well, yes. Or if you assign the IP to the device, if you do static IPs rather than DHCP...

Leo: Right.

Steve: ...then you could assign the IP to a device and prevent it from ever egressing any traffic. Never allow that IP to have outbound traffic. And that would be safe.

Leo: Okay. So it could be on your LAN if you've blocked outbound traffic.

Steve: For its IP. Because it could then never contact headquarters.

Leo: It could do LAN traffic, but not WAN traffic, basically.

Steve: Exactly.

Leo: Okay. That's an interesting approach. So everything could be on the same LAN, just for the IoT devices block WAN egress.

Steve: Right. And actually the way to do that would be to give it its own /8. Normally you're 192.168.0.something. Give it .40 or, well, no, 192.169.99.something. And that way - so then you set up DHCP where you map the MAC address to an IP in that range. So what that's going to do is your IoT devices would all be in the .99. something range. And then you just have one rule on your WAN saying never allow any traffic outbound from .99.anything.

Leo: Right. They can phone some devices on the LAN, but not out to the outside.

Steve: Exactly.

Leo: I would imagine, I wonder if routers aren't going to start building this in, an IoT WAN.

Steve: The fact that they built in Guest with Isolation suggests that they're beginning to get good about it.

Leo: Yeah, yeah.

Steve: Yeah.

Leo: Well, I might have to go through - I have a lot of IT devices. I have more than 100 devices on my network. I'm probably helping Russia right now, Steve.

Steve: I can't even - Leo, if you looked at your raw traffic - now, you have very straight hair now. If you ever came - if we saw you with a perm...

Leo: Curly. Curly curly. I bet - I'm going to have to look and see what Ubiquiti does. Because, I mean, that's one of the advantages of using Ubiquiti is they have a lot of sophisticated security on there. I bet you there's something in there. I have to look. Well, you've certainly opened our eyes to residential proxies. Thank you, Steve. And everything else.

Copyright (c) 2014 by Steve Gibson and Leo Laporte. SOME RIGHTS RESERVED

This work is licensed for the good of the Internet Community under the Creative Commons License v2.5. See the following Web page for details: http://creativecommons.org/licenses/by-nc-sa/2.5/