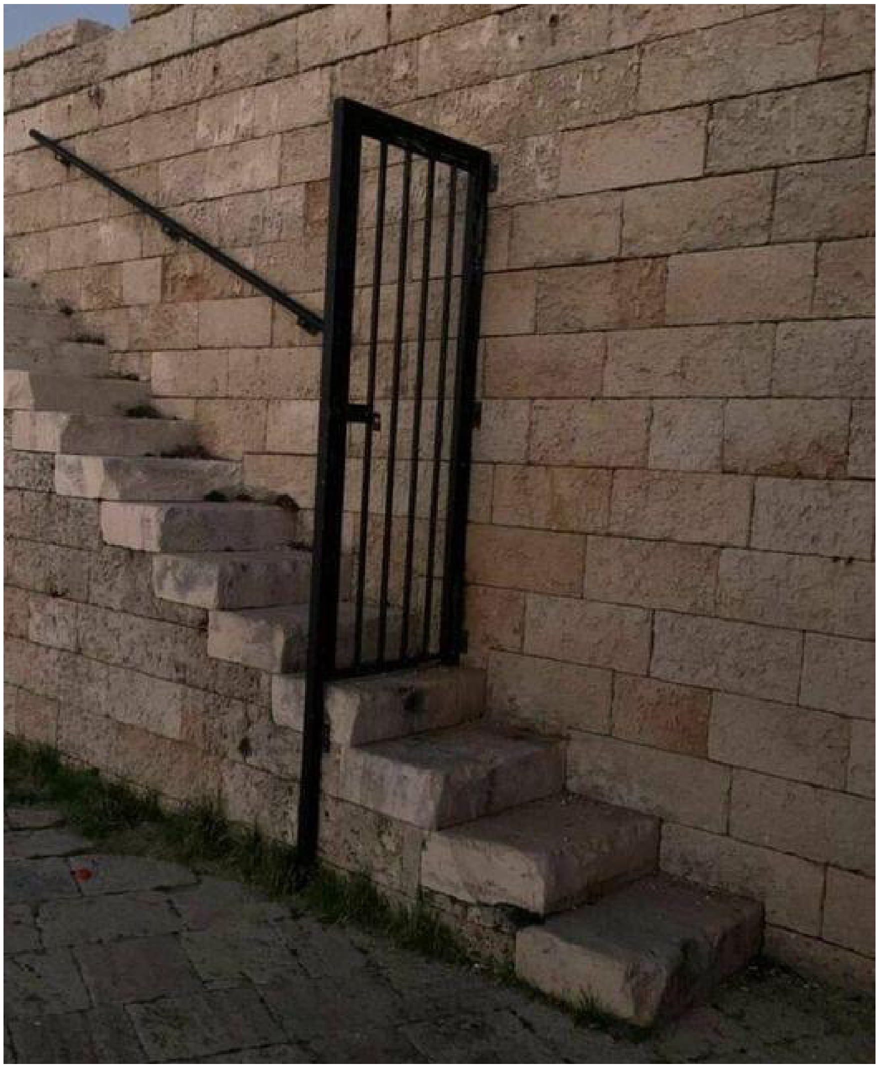
Security Now! #916 - 03-28-23

Microsoft's Email Extortion

This week on Security Now!

In this week's grab bag question collection we wonder: What happened and who cleaned-up during last week's elite 2023 Pwn2Own competition? What happens when GitHub inadvertently exposes their own private SSH RSA key? Are all DDoS for hire site legitimate? And is legitimate ever a word we can apply? Just how bad has the malicious open-source registry package problem become? And how is it that Russia's presidential staff are still using iPhones? After its rocky start in the limelight, how has Zoom's security been faring these past few years? And what benefits can be derived from the sum of two sine waves along a logarithmic curve? What new feature is Microsoft exploring for their already feature-encumbered web browser? And in one of my blessedly rare rants we're then going to learn what new "revenue harvesting" measure Microsoft has just announced which seems deeply ethically wrong to me.

Well… **that**'ll stop'em!



Security News

Synacktiv wins this year's CanSecWest Pwn2Own

Synacktiv’s name came up last week when the firm GRIMM was asked to double-check and verify the result of Synacktiv's forensic reverse-engineering of DJI's drone-controlling software. Since Synacktiv's report could have been seen as highly inflammatory, depending upon how much one believes that we actually have any true security in the first place, it appeared that GRIMM was brought in to obtain the classic “second opinion.”

So, if we didn’t already know that Synacktiv clearly knows their stuff, their breathtaking performance during last week’s 3-day annual Pwn2Own hacking contest, held in Vancouver, Canada, would stand as testimony.

Through the years of this podcast we've been tracking these Pwn2Own competitions, both because they're a lot of fun, and because it never hurts to have an occasional reality check to remind us that whenever skilled hackers take aim at some technology — pretty much any technology — which everyone believes to be secure, that belief is quickly proven to be merely wishful thinking.

There are three Pwn2Own hacking contests that take place throughout the year: one is dedicated to smartphones and IoT devices, and we recently covered that one, another is dedicated to industrial equipment, which we also talked about, and the third is last week’s CanSecWest event that's dedicated to desktops, servers, and smart cars. This competition is widely regarded at the premiere and most prestigious of all hacking contests. And this was not Synacttiv's first Pwn2Own, win since they also took the title two years ago after the competition in 2021. During the multi-category 3-day event, the team successfully demonstrated:

- A heap overflow vulnerability and an out-of-bounds write error in a Bluetooth chipset which allowed them to break into Tesla’s infotainment system and, from there, they were able to gain root access to the rest of the car. That bit is wizardry netted them a cool quarter million dollars ($250,000) and Pwn2Own’s first ever Tier 2 award which is a designation the contest organizers reserve for particularly impactful vulnerabilities and exploits.

- They also demonstrated an attack known as a “TOCTOU” which stands for time-of-check-to-time-of-use. The less fancy term for this is a “race condition” where a time-critical sequence of events can be used to slip past a system's defenses — like query for some status then do something that the designers didn't anticipate rather than waiting for the answer. In this instance they pulled off an attack on Tesla’s Gateway energy management system. They showed how they could then — among other things — open the front trunk or side door of a Tesla Model 3 while the car was in motion. That less than two-minute attack earned the researchers a new Tesla Model 3 (talk about Pwn2Own!) plus a cash reward of $100,000.

- They pulled off a three-bug chain against Oracle's VirtualBox with a host elevation of privilege. That added $80,000 to their rapidly growing winnings.

- The used another TOCTOU bug to escalate privileges on Apple macOS, earning another $40,000.

- By leveraging an incorrect pointer scaling they were able to elevate their privileges on Ubuntu's Desktop Linux to win $30,000.

- And, finally they leveraged a user-after-free flaw against Windows 11 for another $30,000.

So, overall they took home more than half a million dollars ($530,000) and a shiny new Tesla Model 3 earning them the largest award ever raked in by one contestant in Pwn2Own's history.

STARLabs, the runner up, took home a respectable $195,000, which was not bad either.

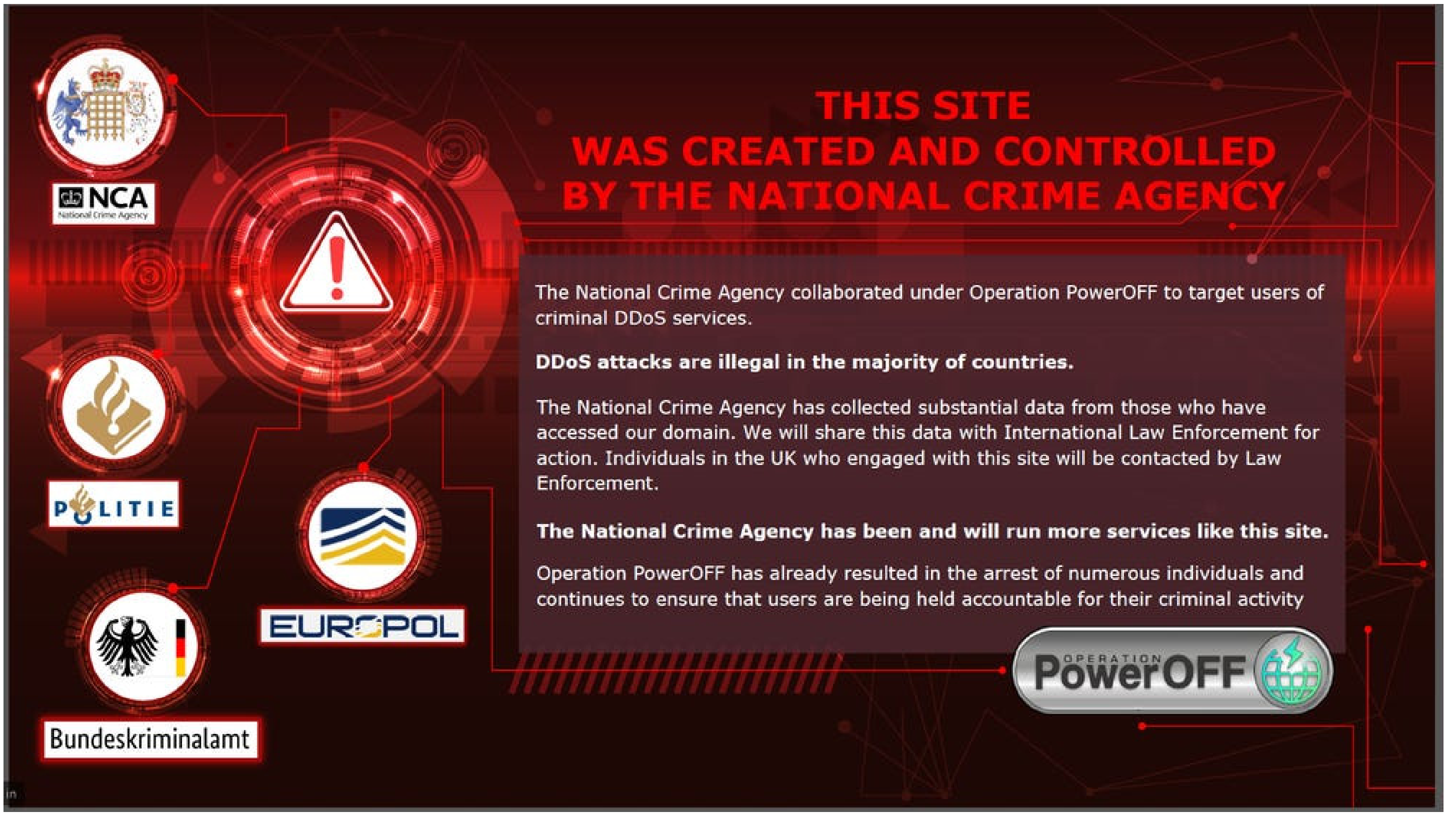
GitHub: Mistakes happen

We were just talking about the benefit of having GitHub repositories continuously scanned for any inadvertent leakage of secret data, such as keys which should never be published. So it was interesting that this just happened to GitHub themselves, causing them to rotate their primary SSH RSA key. Here's what GitHub explained:

> *At approximately 05:00 UTC on March 24, out of an abundance of caution, we replaced our RSA SSH host key used to secure Git operations for GitHub.com. We did this to protect our users from any chance of an adversary impersonating GitHub or eavesdropping on their Git operations over SSH. This key does not grant access to GitHub's infrastructure or customer data. This change only impacts Git operations over SSH using RSA. Web traffic to GitHub.com and HTTPS Git operations are not affected.*
>
> *Only GitHub.com's RSA SSH key was replaced. No change is required for ECDSA or Ed25519 users.*
>
> *This week, we discovered that GitHub.com's RSA SSH private key was briefly exposed in a public GitHub repository. We immediately acted to contain the exposure and began investigating to understand the root cause and impact. We have now completed the key replacement, and users will see the change propagate over the next thirty minutes. Some users may have noticed that the new key was briefly present beginning around 02:30 UTC during preparations for this change.*
>
> *Please note that this issue was not the result of a compromise of any GitHub systems or customer information. Instead, the exposure was the result of what we believe to be an inadvertent publishing of private information. We have no reason to believe that the exposed key was abused and took this action out of an abundance of caution.*
> *What you can do*
>
> *If you are using our ECDSA or Ed25519 keys, you will not notice any change and no action is needed.*

They didn't share anything about how this happened, and there's really no point in them doing so. But it sounded as though they figured it out and I'd imagine that they'll take steps to keep that from happening again.

DDoS for hire... or not!
I love this idea. It makes so much sense. The UK's National Crime Agency says its agents have created several fake DDoS-for-hire services that are up and running today. Such sites can have the dual purpose of catching those in the act of attempting to hire DDoSers as well as frightening others away when the nature of the sting operation is revealed. To serve that second agenda, the National Crime Agency chose to reveal one of its fake DDoS sites last week by replacing the site's earlier homepage with a splash screen announcing the chilling truth:



THIS SITE
WAS CREATED AND CONTROLLED
BY THE NATIONAL CRIME AGENCY

The National Crime Agency collaborated under Operation PowerOFF to target users of criminal DDoS services.

DDoS attacks are illegal in the majority of countries.

The National Crime Agency has collected substantial data from those who have accessed our domain. We will share this data with International Law Enforcement for action. Individuals in the UK who engaged with this site will be contacted by Law Enforcement.

The National Crime Agency has been and will run more services like this site.

Operation PowerOFF has already resulted in the arrest of numerous individuals and continues to ensure that users are being held accountable for their criminal activity

The NCA didn't, wouldn't and shouldn't say how many such sites it's currently running, but did say that it had collected the data of "several thousand people" who have registered on the sites so far. The thing I love about this is that by making clear that DDoS-for-hire sting sites are being operated by law enforcement, a great many users will be scared off and will think better of commissioning a DDoS attack.

144,000 malicious packages published
I saw a statistic that was somewhat sobering. We've been looking at the new challenges facing online open-source repositories which are increasingly being poisoned by a flood of malicious package uploads. These fall under the umbrella of supply-chain attacks. The developer security firm Snyk says it recorded more than 6,800 malicious libraries uploaded on the npm and PyPI portals since the start of this year. So not yet three months. The number that caught my eye was that Snyk said that this recent batch brings the grand total of specifically identified malicious packages to more than 144,000 published to open-source software registries over the past several years since this growing problem was identified.

Huh?
So, the Russian news publication Kommersant reports that the Kremlin's security team has instructed Russia's entire presidential staff to discontinue all use of iPhones by April 1. Kommersant reports that employees were told to get an Android device, either from a Chinese vendor or one running Rostelecom's Aurora OS. And I loved this: The Kremlin officials cited security considerations as being behind their decision, claiming that iPhones were "more susceptible to hacking and espionage by Western experts compared to other smartphones." Huh? That doesn't correspond to anything we know.

This is another of the recent examples we've looked at where, in an environment of increasing mistrust and hostility, it really doesn't make any sense for anyone to be using a closed device sourced from any entity on the other side of the dispute – and Apple's iPhones are certainly far more closed than Android devices. So, yeah, if Russia is planning a long-term split from the West, then discontinuing all use of Western-sourced tech is the only sane long-term strategy.

Zooming right along…
In our industry's apparently eternal quest to rid ourselves of mistakes made in the creation of software, one of the more effective strategies that has been found is the idea of paying good guys to find and report those flaws before bad guys can find them and use them against us. Thus bug bounty programs have become a mainstay. The COVID-driven work from home boom, quickly put a lesser known video conferencing system, Zoom, on the map. But not everything went well from the start. As we have seen time and time again, many more bugs exist than are known. So Zoom's pre-celebrity confidence in its software was quickly shaken when bad guys began looking more closely at it than ever before and discovered all sorts of ways that its benefits could be subverted. As we covered at the time, this came as quite a shock to Zoom's management and they did stumble a bit out of the gate. But to their credit they quickly hired some experienced right thinking true security experts and the establishment of a functioning bug bounty program was near the top of the list.
So how has that been going? Here's how Roy Davis, Zoom's Security Manager described this effort in a blog posting last week:

> *In security, it's all about who gets there first. We race to identify bugs and issues before the bad guys do, so we tap the ethical hacking community to help us get ahead.*
>
> *We source this help through our Zoom Bug Bounty program, which lets us connect with and engage expert researchers that help us proactively mitigate risk and create a safer environment for our customers. And we've accomplished a lot as a community in the past year. Here's a look:*
>
> *We test our infrastructure every day at Zoom, but we know we're not immune to edge-case vulnerabilities. So, we call in backup — the ethical hacker community can sometimes detect bugs that may only be discovered in certain circumstances.*
>
> *That's why our bug bounty program focuses on recruiting skilled, effective researchers. In 2022, we sent additional invitations to researchers to join our HackerOne program with a focus*

> *on attracting active security talent. We also like to go beyond our program to find talent, so we tapped into the community via industry events like H1-702. (A HackerOne event)*
>
> *These researchers work hard to help us, so we strive to celebrate successful report submissions accordingly. In the fiscal year 2023, we awarded **$3.9 million** in bounties to hundreds of researchers and over **$7 million** to date since the program began.*
>
> *Beyond identifying vulnerabilities, outside researchers' support has helped us make other forms of progress at Zoom. We used these reports to demonstrate items that needed attention, flag root-level causes for issues, create better cross-functional alignment, and find potential threats before they become a problem. As a result, our time-to-resolution for bug bounty reports has significantly improved over the past two years.*
>
> *At the start of this year, we restructured our team and developed updates for the program for FY24. We evaluated the researchers currently in our program to make sure everyone is active and contributing. We want to put the right foot forward in the new year, and that all starts by working with high-caliber, effective researchers.*
>
> *Zoom's Bug Bounty program is also implementing a brand new vulnerability impact scoring system to help researchers do their best work yet. While we will continue to use the industry standard Common Vulnerability Scoring System (CVSS) to score reports, we're evolving our program to add a companion scoring system called the Vulnerability Impact Scoring System (VISS) that analyzes 13 different aspects of impact for each vulnerability reported as they relate to the Zoom infrastructure, technology, and security of customer data. With the implementation of VISS, Bug Bounty can focus more on measuring responsibly demonstrated impact, rather than the theoretical possibility of exploitation.*
> *The road ahead*
>
> *As the Zoom Bug Bounty program has grown over the past year, we're continuing to evolve and mature our processes, bounty awards, and testing scope. We're very excited to see the impact of our new scoring system and all the good our researchers can do in 2023.*
>
> *If you're interested in helping to make Zoom more secure, email your HackerOne profile name to bugbounty@zoom.us or visit the Zoom careers page to review the open positions within the Trust and Security teams. Happy hacking!*

This is what being proactive about security looks like. Yes, as we all know, since we chronicled those early failures, Zoom was initially caught flat footed when their platform took off. But today Zoom's security team is **actively** managing their bug bounty program. And that is clearly making a big difference. They're not simply passively listed over at HackerOne and claiming for the sake of a bullet point on a presentation slide that "oh yes, we offer bug bounties." No. They're serious about tightening up their platform.

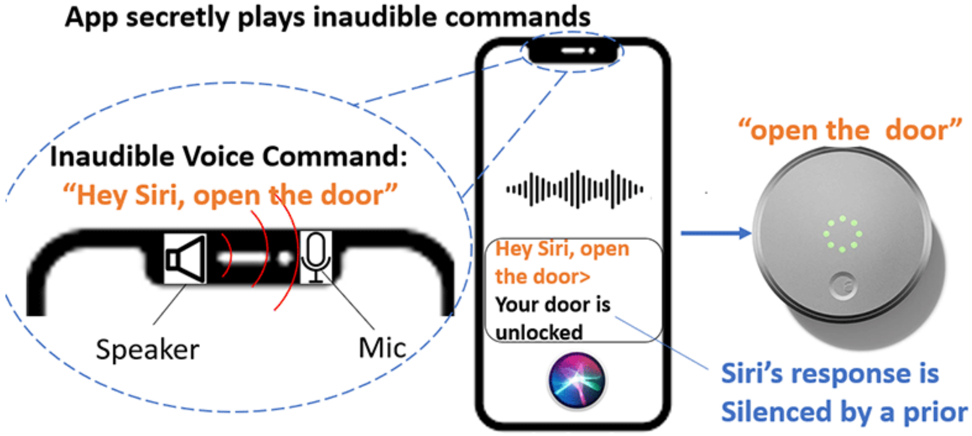
That H1-702 event that Roy referred to? That was a multi-day HackerOne event held in Las Vegas, Nevada early last August and Zoom was one of two corporate sponsors of the live hacking event on August 4th during which more 100 security professionals (around 70 in-person and 40 virtual) from 29 countries hacked the Zoom web and desktop client, APIs, Zooms Marketplace apps, and any of the binaries that Zoom distributes. Five individual awards were distributed and overall Zoom paid roughly $480,000 in bounties at the event. They said that they feel this is a reflection of the importance of this industry best practice – meaning paying bounties for responsibly reported bug discoveries. They've come a long way from where they started when they first popped onto our radar in less than stellar fashion. Bravo, Zoom!

I NUIT!

As in "I knew it!" — Actually, it's NUIT which in French means "nighttime," NUIT is an acronym for "Near-Ultrasound Inaudible Trojan." And so, yes, some clever researchers are again going to entertain us with their out of the box thinking.

Researchers from the University of Texas at San Antonio and the University of Colorado at Colorado Springs recently published a paper for presentation during the USENIX Security 2023 conference being held in April next month. It demonstrates a novel inaudible voice Trojan attack which exploits vulnerabilities of smart device microphones and voice assistants — like Siri, Google Assistant, Alexa, Cortana and so on.

The researchers used their Near-Ultrasound Inaudible Trojan to attack different types of smart devices, bridging from smart phones to smart home devices. The results of their demonstrations show that NUIT is effective in maliciously controlling the voice interfaces of popular tech products and that those tech products which are currently on the market have vulnerabilities.


App secretly plays inaudible commands
Inaudible Voice Command:
"Hey Siri, open the door"
Speaker
Mic
Hey Siri, open the door>
Your door is unlocked
"open the door"
Siri's response is Silenced by a prior

The NUIT attack takes advantage of the fact that digital assistants use microphones which accurately pickup sounds that are inaudible to the human ear. NUIT plays sounds in the near-ultrasound frequency range from 16 to 20kHz which enables it to give voice commands to both close and more remote smart devices.

Their research demonstrated that NUIT-style near-ultrasound commands can be embedded pretty much anywhere. An attacker could direct a victim to click a link to a website or a YouTube video which would then play the inaudible voice commands. The researchers demonstrated that NUITs also work when playing from one phone which controls another, over Zoom calls, playing on a phone to control a smart speaker or other IOT device, or even embedded into files that have additional background music.

Once they have unauthorized access to a device, hackers can send inaudible action commands to reduce a device's volume and prevent a voice assistant's response from being heard by the user before proceeding with further attacks. I was unable to find their full research paper online and the USENIX conference, as I mentioned, isn't until next month, so some puzzles remain. In some summary coverage published by their universities they're quoted saying that to wage a successful attack against voice assistant devices, the length of malicious commands must be shorter than 0.77 seconds but we won't know why until their formal paper is published.

They did add that the vulnerability is created due to the nonlinearity of the microphone design, which the manufacturer would need to address. And the researchers said that out of the 17 smart devices they tested, Apple's Siri devices alone needed to capture and reuse their user's voice while other voice assistant devices were activated by using any voice or a robot voice. They also pointed out that the attack could be surreptitious because it was possible to silence Siri's response since iPhones maintain separate volumes for Siri and non-Siri output.

Users of voice assistants have all experienced odd triggers of the system when it did not appear that the system was being addressed. These may be the result of their microphones hearing and responding to a much wider range of frequencies than humans do. And while there's still a lot that we don't know about the mechanism of this attack, there's an interesting opportunity for a bit of science and math conjecture here.

There was the comment made that the attack is due to non-linearities in the operation of these microphones. And that provided the clue. That almost certainly means that the instantaneous response to air pressure sound waves is not linear. If the response was non-linear at normal operating volume the result would be unacceptable distortion. But the non-linearity is likely to be extreme at very low volume levels where that doesn't matter.

Any time you have a non-linear response, the **addition** of two inputs along that non-linear response curve is wonderfully turned into **multiplication**. Although this may initially be counter-intuitive, this is the principle of logarithms and it's the way a slide rule, which adds linear lengths, is able to produce multiplication. The scales of a slide rule are logarithmically non-linear. So when you're adding linear lengths on a slide rule you're performing multiplication.

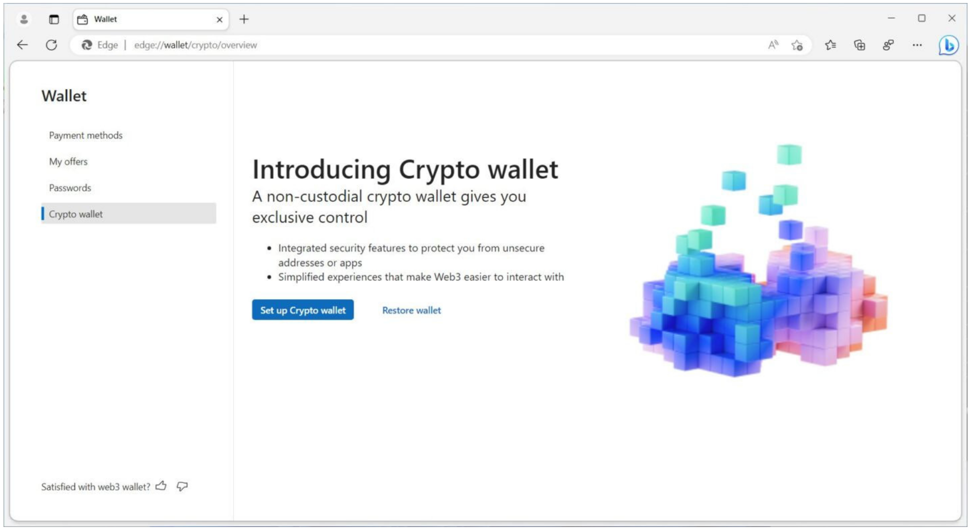
This means that if we had two sine waves at very low volume, their summation by the device's microphone having a nonlinear response at low volume, would have the effect of multiplying their values in real time. So next we add one of my favorite trigonometric identities which states that the product of two sine waves is equal to the sum and the difference of their frequencies.

Leo, you have your amateur radio operator's license, so you know of this as heterodyning. In radio, heterodyning is the way a radio's local oscillator is able to bring a radio frequency signal down to audible frequency range. What we hear is the difference between the two frequencies, neither of which are audible... and that's exactly what's happening here.

The researchers are generating a pair of near-ultrasonic frequencies whose difference is the voice signal they are using to control other devices. We don't hear anything. But the microphones in those devices, which are always straining to hear our commands, believe that they're hearing our voice because inaudible sine waves are being made to heterodyne.

Edge gets Crypto

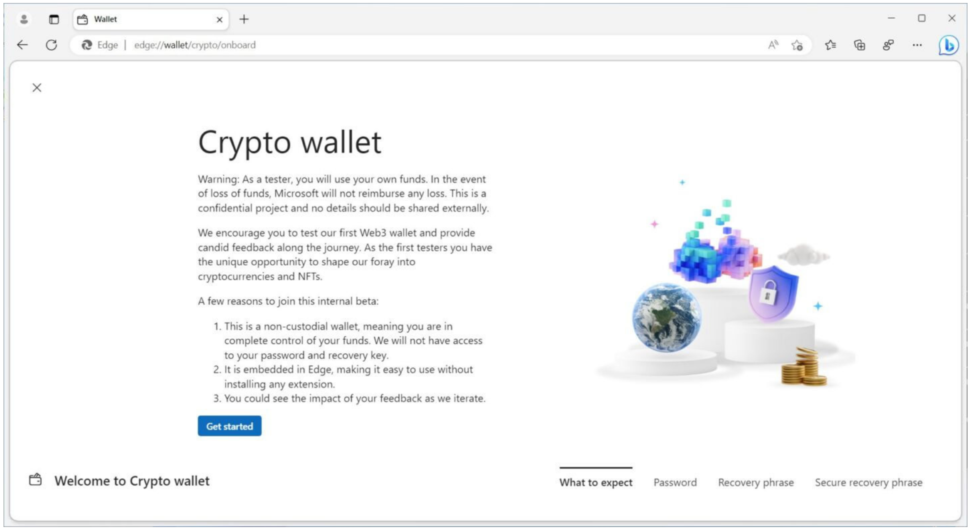
The news is that Microsoft has started testing a cryptocurrency wallet which they are planning to build into their — let's just say, "increasingly versatile" — Edge browser. I liked ArsTechnica's take on this. Their headline read: *"Microsoft is testing a built-in cryptocurrency wallet for the Edge browser"* with the subhead: *"Crypto wallet would join coupons, cash back, and "buy now, pay later" add-ons."*



Introducing Crypto wallet

A non-custodial crypto wallet gives you exclusive control

- Integrated security features to protect you from unsecure addresses or apps
- Simplified experiences that make Web3 easier to interact with



Crypto wallet

Warning: As a tester, you will use your own funds. In the event of loss of funds, Microsoft will not reimburse any loss. This is a confidential project and no details should be shared externally.

We encourage you to test our first Web3 wallet and provide candid feedback along the journey. As the first testers you have the unique opportunity to shape our foray into cryptocurrencies and NFTs.

A few reasons to join this internal beta:

1. This is a non-custodial wallet, meaning you are in complete control of your funds. We will not have access to your password and recovery key.
2. It is embedded in Edge, making it easy to use without installing any extension.
3. You could see the impact of your feedback as we iterate.

Andrew Cunningham is Ars Technica's Senior Technology Reporter whose take on this is, I think, spot-on. Here's how Andrew explained and characterized this new find:

> *Microsoft appears to be testing a built-in cryptocurrency wallet for Edge, according to screenshots pulled from a beta build of the browser. The feature, which the screenshots say is strictly for internal testing, was unearthed by Twitter user @thebookisclosed, who has a history of digging up present-but-disabled features in everything from new Windows 11 builds to ancient Windows Vista betas.*
>
> *This is only one of many money and shopping-related features that Microsoft has bolted onto*

> *Edge since it was reborn as a Chromium-based browser a few years ago. In late 2021, the company faced backlash after adding a "buy now, pay later" short-term financing feature to Edge. And as an Edge user, the first thing I do in a new Windows install is disable the endless coupon code, price comparison, and cash-back pop-ups generated by Shopping in Microsoft Edge (many settings automatically sync between Edge browsers when you sign in with a Microsoft account; the default search engine and all of these shopping add-ons need to be changed manually every time).*
>
> *According to the screenshots, the crypto wallet is "embedded in Edge, making it easy to use without installing any extension," and it can handle multiple types of cryptocurrency. It will also record transactions and the value of your individual currencies as they fluctuate. An "explore" tab offers news stories relevant to cryptocurrency, and an "assets" tab will let you stare lovingly at your NFTs. The wallet is "non-custodial" (also called "self-custodial"), meaning that you have sole ownership of and responsibility for the passwords and recovery keys that allow access to your funds. Microsoft won't be able to let you back in if you lose your credentials.*
>
> *Whether you find these kinds of add-ons useful, annoying, or predatory is a matter of perspective. Given the prevalence of crypto scams, there may be some value in having a "trustworthy" built-in option that doesn't require the installation of dodgy third-party extensions. But the feature could also encourage casually interested users to begin exploring the world of cryptocurrency, which is, again, rife with scams.*
>
> *It's also yet another example of Microsoft building a not strictly browsing-related feature into its web browser. Many of these features can be disabled, and competing browsers like Chrome and Firefox all attempt to add value and earn money by building in access to new niche features and third-party services. But Microsoft's moves can still have an outsize impact that deserves extra scrutiny—Edge is an installed-by-default, non-removable component of every Windows 10 and Windows 11 PC, and the operating system pushes you to switch to Edge with some regularity. And once in Edge, the browser pushes you to use Bing and other Microsoft services.*
>
> *Microsoft may not ship the crypto wallet to Edge users—the company regularly tests features in Edge, Windows, and its other software that never end up making it into the general-release versions. We've contacted Microsoft for more information and will update if we receive a response.*

Not long ago I tried to use Bing. I had become annoyed with Chrome because I noticed that every time I opened it the fan on my little Intel NUC would spin up to dissipate the heat that Chrome was, for some reason, causing my system to produce. And this was with no tabs loaded. Just Chrome itself. So it had become bloatware. I didn’t know whether Bing might be any better, but I need left side tabs instead of top tabs, so Bing’s built-in support for that feature drew me in. I figured that being first and foremost a Chromium-based browser, I’d at least get good compatibility. But then I found that some web pages would not open or display in Bing. So back to Firefox I went where I am once again completely happy. If Microsoft decides to embed a cryptocurrency wallet into Bing, I hope it works better than their eMail solutions have.

Microsoft’s Email Extortion

So, I should start out by noting that it’s been quite a while since the listeners of this podcast have heard me get really upset about anything. It doesn’t happen very often, and I can’t recall the last time it happened, but neither does it feel entirely unique. When I dwell on this one I’m sure my blood pressure rises because I have a real problem with injustice and bullying. Especially when technology is used to perpetrate it. Someone at Microsoft has had a **very** bad idea.

When I first encountered this, yesterday, I did a double-take. I thought that I must have misunderstood what Microsoft meant. But unfortunately, no. Microsoft has formally announced that they are going to begin blocking incoming eMail to their Exchange Online cloud instances which includes all of Office 365 and Outlook.com, if that incoming eMail originates from other private so-called on premises Exchange Servers which, while they may be functioning just fine, are nevertheless, past their end-of-support life. Wow.

That’s right. They are saying that they are going to begin blocking incoming eMail from older version instances of their own Exchange Server software. No one who purchased Exchange Server 2007, 2010 or 2013 was told at the time — in fact they’ve not been told until now — that in the future, the software they purchased, paid for and have continued to happily use, would become less useful to them because Microsoft’s various online eMail services were going to unilaterally begin refusing to accept eMail from those perfectly functioning servers.

And, yes, I use the term “perfectly functioning” in the context of Exchange Server with a bit of tongue in cheek, because, after all, it still is Exchange Server. But many hundreds of thousands of instances of it are still functioning, and everyone else in the world will be able to receive the eMail they send, except for Microsoft's services, because Microsoft has apparently decided to punish their previous customers and extort them for additional licensing revenue.

https://techcommunity.microsoft.com/t5/exchange-team-blog/throttling-and-blocking-email-from-persistently-vulnerable/ba-p/3762078

Now, interestingly, the headline on Microsoft's announcement doesn’t quite fess up to this fully. It reads: *“Throttling and Blocking Email from Persistently Vulnerable Exchange Servers to Exchange Online.”* But they define the term “persistently vulnerable” as meaning *“Servers that are unsupported or remain unpatched.”* And just wait until you hear their rationale for doing this. My breath is still a bit taken away by this. I want to repeat my summation so that you don’t need to hit rewind or replay to be sure you heard it right: Microsoft is essentially saying that they are going to use their market dominance in online eMail services to force their previous software customers to upgrade their own instances of Exchange Server by refusing to accept eMail sent by such servers as a means of extorting additional licensing fees from those prior customers. Microsoft is going to effectively begin reducing the functionality of their previous Exchange Servers by refusing to accept their eMail unless and until the licenses for those Exchange Servers are renewed and their software is updated.

I cannot think of a precedent for this in our industry. So I suppose that means that this is an unprecedented action. Here's how Microsoft couched this extortion. They wrote:

> *As we continue to enhance the security of our cloud, we are going to address the problem of email sent to Exchange Online from unsupported and unpatched Exchange servers. There are many risks associated with running unsupported or unpatched software, but by far the biggest risk is security. Once a version of Exchange Server is no longer supported, it no longer receives security updates; thus, any vulnerabilities discovered after support has ended don't get fixed. There are similar risks associated with running software that is not patched for known vulnerabilities. Once a security update is released, malicious actors will reverse-engineer the update to get a better understanding of how to exploit the vulnerability on unpatched servers.*

Yep. As we know, all of that's true. But next comes the pivotal paragraph. The fundamental flaw in the logic upon which this entire extortion effort rests. Microsoft continues...

> *Microsoft uses the Zero Trust security model for its cloud services, which requires connecting devices and servers to be provably healthy and managed. Servers that are unsupported or remain unpatched are persistently vulnerable and cannot be trusted, and therefore email messages sent from them cannot be trusted. Persistently vulnerable servers significantly increase the risk of security breaches, malware, hacking, data exfiltration, and other attacks.*

Repeating that crucial line: *"Servers that are unsupported or remain unpatched are persistently vulnerable and cannot be trusted, and therefore email messages sent from them cannot be trusted."* This is what's commonly known as a load of crap. It's important we pause here for a minute because, as I said, it is upon this fundamental logical fallacy that Microsoft is hanging their entire campaign.

We know that the truth is that, yes, through the years Microsoft's Exchange Server has had a particularly difficult relationship with security. In short, it's pretty much been an utter disaster. And not just for a while. It's inexplicable. eMail is a trivial protocol. But Microsoft has managed to make it deadly.

On January 3rd of this year, BleepingComputer's headline read: *"Over 60,000 Exchange servers vulnerable to ProxyNotShel"* And few months earlier on October 27th, Wired's headline (and I'm not making this up) was *"Your Microsoft Exchange Server Is a Security Liability"* with the subhead: *"Endless vulnerabilities. Widespread hacking campaigns. Slow and technically tough patching. It's time to say goodbye to on-premise Exchange."* As an aside, I'll remind everyone of just how conscientious Microsoft has been about Exchange's security. In that article, describing the constant struggle over Exchange vulnerabilities, Wired wrote:

> *The latest reminder of that struggle arrived earlier this week, when Taiwanese security researcher Orange Tsai published a blog post laying out the details of a security vulnerability in Microsoft Exchange. Tsai warned Microsoft about this vulnerability as early as June of 2021, and while the company responded by releasing some partial fixes, it took Microsoft 14 months to fully resolve the underlying security problem. Tsai had earlier reported a related vulnerability in Exchange that was massively exploited by a group of Chinese state-sponsored*

> *hackers known as Hafnium, which last year penetrated more than 30,000 targets by some counts. Yet according to the timeline described in Tsai's post this week, Microsoft repeatedly delayed fixing the newer variation of that same vulnerability, assuring Tsai no fewer than four times that it would patch the bug before pushing off a full patch for months longer. When Microsoft finally released a fix, Tsai wrote, it still required manual activation and lacked any documentation for four more months.*
>
> *Meanwhile, another pair of actively exploited vulnerabilities in Exchange that were revealed last month still remain unpatched after researchers showed that Microsoft's initial attempts to fix the flaws had failed. Those vulnerabilities were just the latest in a years-long pattern of security bugs in Exchange's code. And even when Microsoft does release Exchange patches, they're often not widely implemented, due to the time-consuming technical process of installing them.*

So let's first just be really clear here: Any security problems that Exchange Server has are directly Microsoft's fault. So, now they're saying that they're **afraid of receiving** any eMail that previous versions of their software might send to their current versions. It's true that the poor users of those previous versions are likely risking life and limb to continue running older and out-of-date versions of Exchange. But that's their choice (or it used to be), and it's **they** who are being put in danger by running Microsoft's perennially insecure offering, not people who receive the eMail it sends. That's utter nonsense. But it's the utter nonsense upon which, as I noted, this entire campaign rests.

So here's how Microsoft continues:

> *We've said many times that it is critical for customers to protect their Exchange servers by staying current with updates and by taking other actions to further strengthen the security of their environment. Many customers have taken action to protect their environment, but there are still many Exchange servers that are* ***out of support*** *or significantly behind on updates.*

So Microsoft has coined a new term of art for this which will be entering our industry's lexicon. They are calling what they will be employing their *"Transport-based Enforcement System."*

> *To address this problem, we are enabling a transport-based enforcement system in Exchange Online that has three primary functions: reporting, throttling, and blocking. The system is designed to alert an admin about unsupported or unpatched Exchange servers in their on-premises environment that need remediation (upgrading or patching). The system also has throttling and blocking capabilities, so if a server is not remediated, mail flow* ***from*** *that server* ***will*** *be throttled (delayed) and eventually blocked.*
>
> *We don't want to delay or block legitimate email, but we do want to reduce the risk of malicious email entering Exchange Online by putting in place safeguards and standards for email entering our cloud service.* ***We also want to get the attention of customers who have unsupported or unpatched Exchange servers and encourage them to secure their on-premises environments.***

That's right. "Encouragement" is what we call it here at Microsoft when your eMail server's outbound mail starts being refused, not due to any misbehavior on its part, which would be your problem in any event, **but only because** you have chosen to continue using an older version of Exchange Server – which you paid for – whose eMail output we have since decided we no longer want to accept. **Oh!!** But if you purchase a new license then all will be forgiven and we'll happily receive your eMail.

They also explain about the actions they'll be taking:

> *If a server is not remediated after a period of time, Exchange Online will begin to throttle messages from it. In this case, Exchange Online will issue a retriable SMTP 450 error to the sending server which will cause the sending server to queue and retry the message later, resulting in delayed delivery of messages. In this case, the sending server will automatically try to re-send the message.*
>
> *The throttling duration will increase progressively over time. Progressive throttling over multiple days is designed to drive admin awareness and give them time to remediate the server. However, if the admin does not remediate the server within 30 days after throttling begins, enforcement will progress to the point where email will be blocked.*
>
> *If throttling does not cause an admin to remediate the server, then after a period of time, email from that server will be blocked. Exchange Online will issue a permanent SMTP 550 error to the sender, which triggers a non-delivery report (NDR) to the sender. In this case, the sender will need to re-send the message.*
>
> *We're intentionally taking a progressive enforcement approach which gradually increases throttling over time, and then introduces blocking in gradually increasing stages culminating in blocking 100% of all non-compliant traffic.*
>
> *Enforcement actions will escalate over time (e.g., increase throttling, add blocking, increase blocking, full blocking) until the server is remediated: either removed from service (for versions beyond end of life), or updated (for supported versions with available updates).*

The announcement of this then contains an FAQ to clarify, in Q&A format, what that means. The answers to a few of the questions they ask themselves are a bit chilling...

> ***What is a persistently vulnerable Exchange server?***
>
> Any Exchange server that has reached end of life (e.g., Exchange 2007, Exchange 2010, and soon, Exchange 2013), or remains unpatched for known vulnerabilities. For example, Exchange 2016 and Exchange 2019 servers that are significantly behind on security updates are considered persistently vulnerable.
>
> ***Is Microsoft blocking email from on-premises Exchange servers to get customers to***

move to the cloud?

No. Our goal is to help customers secure their environment, wherever they choose to run Exchange. The enforcement system is designed to alert admins about security risks in their environment, and to protect Exchange Online recipients from potentially malicious messages sent from persistently vulnerable Exchange servers.

Why is Microsoft only taking this action against its own customers; customers who have paid for Exchange Server and Windows Server licenses?

We are always looking for ways to improve the security of our cloud and to help our on-premises customers stay protected. This effort helps protect our on-premises customers by alerting them to potentially significant security risks in their environment. We are initially focusing on email servers we can readily identify as being persistently vulnerable, but we will block all potentially malicious mail flow that we can.

Will Microsoft enable the transport-based enforcement system for other servers and applications that send email to Exchange Online?

We are always looking for ways to improve the security of our cloud and to help our on-premises customers stay protected. We are initially focusing on email servers we can readily identify as being persistently vulnerable, but we will block all potentially malicious mail flow that we can.

If my Exchange Server build is current, but the underlying Windows operating system is out of date, will my server be affected by the enforcement system?

No. The enforcement system looks only at Exchange Server version information. But it is just as important to keep Windows and all other applications up-to-date, and we recommend customers do that.

Delaying and possibly blocking emails sent to Exchange Online seems harsh and could negatively affect my business. Can't Microsoft take a different approach to this?

Microsoft is taking this action because of the urgent and increasing security risks to customers that choose to run unsupported or unpatched software. Over the last few years, we have seen a significant increase in the frequency of attacks against Exchange servers. We have done (and will continue to do) everything we can to protect Exchange servers but unfortunately, there are a significant number of organizations that don't install updates or are far behind on updates, and are therefore putting themselves, their data, as well as the organizations that receive email from them, at risk. We can't reach out directly to admins that run vulnerable Exchange servers, so we are using activity from their servers to try to get their attention. Our goal is to raise the security profile of the Exchange ecosystem.

Why are you starting only with Exchange 2007 servers, when Exchange 2010 is also beyond end of life and Exchange 2013 will be beyond end of life when the

enforcement system is enabled?

Starting with this narrow scope of Exchange servers lets us safely exercise, test, and tune the enforcement system before we expand its use to a broader set of servers. Additionally, as Exchange 2007 is the most out-of-date hybrid version, it doesn't include many of the core security features and enhancements in later versions. Restricting the most potentially vulnerable and unsafe server version first makes sense.

Does this mean that my Exchange Online organization might not receive email sent by a 3rd party company that runs an old or unpatched version of Exchange Server?

Possibly. The transport-based enforcement system initially applies only to email sent from Exchange 2007 servers to Exchange Online over an inbound connector type of OnPremises. The system does not yet apply to email sent to your organization by companies that do not use an OnPremises type of connector. Our goals are to reduce the risk of malicious email entering Exchange Online by putting in place safeguards and standards for email entering the service and to notify on-premises admins that the Exchange server their organization uses needs remediating.

How does Microsoft know what version of Exchange I am running? Does Microsoft have access to my servers?

No, Microsoft does not have any access to your on-premises servers. The enforcement system is based on email activity (e.g., when the on-premises Exchange Server connects to Exchange Online to deliver email).

I'm not the only person to react negatively to this posting. There has been some interesting public dialog in the replies online.

An Occasional Contributor named Mike Crowley on Mar 23rd writes:
Wow. Perhaps not the intent, but this means if customers don't continue to pay for Exchange, they cannot send mail to Office 365 users.

Scott Schnoll, a Microsoft representative replies:
@Mike Crowley, not sure how you got that idea, but there is no requirement for anyone to continue to pay for Exchange in order to send mail to Office 365 users. There is, however, a requirement to use servers that are provably healthy and managed for sending mail to Exchange Online (and Office 365 users).

(Of course, that's not true, since Microsoft is only refusing eMail from their own eMail servers.)

Another occasional visitor whose handle is "essentialexch" asks:
Are you blocking old versions of sendmail? KerioMail? hMail? postfix? OpenXchange? IceWarp? Notes? The list goes on and on. If you aren't blocking anything but Exchange, then this is definitely a biased and prejudicial decision. (And I say this with no clients running versions of exchange so old as to be affected by this decision.) The conspiracy minded part of me thinks

there is likely something else going on.

Microsoft’s representative replies:
@essentialexch, we are initially focusing on email servers we can readily identify as being persistently vulnerable, but we will continue to look for ways to block all potentially malicious mail flow that we can. But we want to be careful, deliberate, predictable, and transparent, so we're specifically starting with a set of servers belonging to customers we can identify and notify, and with whom we have a relationship. But please keep in mind, that our goal is not to throttle or block anyone's email, but rather to get senders of email to Exchange Online to deal with a serious security issue in their environment that can directly affect others.

So now that we have a good understanding of where we are, where exactly are we?

The world is apparently full of past end-of-life and non-updated Exchange Server instances that are, day in and day out, working perfectly well for their users who once purchased the software from Microsoft and who have felt no need or reason to purchase new licenses. The server is online and working for them the way they want it to without problems; and for whatever reason they have chosen not to upgrade their software. Is it not their choice, and their right, to run the software that they have purchased as long as they wish and as long as it is serving their needs? That software was not a subscription. It did not come with an expiration date. Its license’s clear assertion was that they would be able to use it for as long as they wished.

The idea that eMail originating from out of date and unsupported eMail servers is itself inherently dangerous, is utter nonsense. There’s no awareness out in the world that eMail from certain versions of this or that eMail software is known to be dangerous. Microsoft just invented that out of thin air to suit their commercial purpose. To the extent that there **is** any danger, it’s to the organization running such software, not to those receiving its eMail output. This is the shaky premise upon which Microsoft’s policy rests, and it doesn’t hold up. But that doesn’t matter because Microsoft is able to claim it, and will soon begin enforcing it.

Those perfectly good servers have been tirelessly working, apparently without problems, for many years. But now that’s about to change, not because they suddenly represent any actual danger, **but because Microsoft figured out how they can extort some new revenue from their old customers.**

Because most organizations will take the path of least resistance – which will be to upgrade – Microsoft surely knows that the net effect of this will be to generate additional revenue for itself. Those wayward users have fallen off the Microsoft gravy train and it’s high time that they be brought back into the fold. What better way than to create a new inability for those old servers to send mail to Office 365 and Outlook.com users whom Microsoft already utterly controls? It’s diabolically brilliant. And so that it doesn’t appear to be the bald faced revenue extortion scheme that it is, couch the whole thing in their need to protect their paying customers from those evil eMail messages being generated by their own presumably now highly dangerous software.

And finally, it’s worth noting that this is not a one-time event. No one using any Microsoft Exchange Server software will ever again be able to fail to keep it updated, nor to avoid the purchase of future licenses – forever. I celebrate that idea from here forward, since keeping software updated, especially Exchange Server, is a good thing and since new licensors will be aware that they will never again have any choice other than to be paying Microsoft forever under whatever terms and conditions Microsoft may choose.

The problem I have is with Microsoft's effectively unilateral revocation of the open-ended licenses they **previously** sold. That's a clear violation of trust and that's never going to be okay even though it's going to happen.