SECURITY NOW!

Transcript of Episode #731

DeepFakes

Description: This week we look at a forced two-day recess of all schools in Flagstaff, Arizona; the case of a ransomware operator being too greedy; Apple's controversial response to Google's posting last week about the watering hole attacks; Zerodium's new payout schedule and what it might mean; the final full public disclosure of BlueKeep exploitation code; some potentially serious flaws found and fixed in PHP that may require our listener's attention; some SQRL news, miscellany, and closing-the-loop feedback from a listener. Then we take our first look on this podcast into the growing problem and threat of "Deepfake" media content.

High quality (64 kbps) mp3 audio file URL: http://media.GRC.com/sn/SN-731.mp3
Quarter size (16 kbps) mp3 audio file URL: http://media.GRC.com/sn/sn-731-lq.mp3

SHOW TEASE: It's time for Security Now!. Steve Gibson is here. Yes, we're recording a little early to make room for the Apple event on Tuesday. But don't worry. We've got a great show planned for you. Steve will give you a rundown of some of the latest ransomware wins and losses, including one ransomware author that decided to charge way too much. We'll also talk about deepfakes, how to tell if something's really real. We're quickly headed to an era where not everything you see can be believed. It's all coming up on Security Now!.

Leo Laporte: This is Security Now! with Steve Gibson, Episode 731, recorded Monday, September 9th, 2019: Deepfakes.

It's time for Security Now!, the show where we talk about the latest security and privacy and all that jazz. Steve Gibson is here. He's the man in charge at GRC.com. Hi, Steve.

Steve Gibson: Hey, Leo. I'm looking at one screen where I'm normally looking at you, which says: Working on updates, 23%.

Leo: Oh. It's not even Patch Tuesday. That's tomorrow.

Steve: Yeah, well, it was, you know, I'd been deferring, right, updates because I only use this particular Windows 10 machine for you. It is reserved. It's my Leo computer.

Leo: Thank you.

Steve: And so several hours ago I thought, because it turned on, and it said, oh, you've got updates, you know. And I thought, okay. I looked at the clock. Lots of time. No problem.

Leo: Yeah.

Steve: Uh-huh.

Leo: You know, I can install an entire Linux OS in a tenth of the time that it takes to update Windows. It's crazy.

Steve: You could probably - you could write an operating system in Lisp...

Leo: I probably could.

Steve: ...in the time it takes to install updates.

Leo: Well, I do see - I just ran my update thing. Maybe you're getting 1903, which is kind of ironic because 19H2 is going to be out sometime.

Steve: What I'm getting is dizzy from looking at these little dots running around in circles. It's like...

Leo: And then in a few minutes it'll remind you, all your files, exactly where you left them, and getting things set up.

Steve: So let's not pay attention to the fact that I moved from 21 to 23, and it only took half an hour to make that 2% jump. We're doing odd percentages for some reason. I don't know why. And actually we have an odd podcast. It's 731, and we're recording it on the 9th; but it's technically Security Now! for September 10th because tomorrow is a big Apple announcement day. And I will be glued to the TWiT channel, following...

Leo: Will you really?

Steve: Oh, yeah. I love...

Leo: You still care.

Steve: It's like MSTK, whatever that dumb...

Leo: 3000? 2000?

Steve: Yeah, right, right, where the little shadows were down in the front, making wisecracks at the movie.

Leo: Yeah. That's what we do.

Steve: That's like you guys. That's watching TWiT during an Apple thing.

Leo: And it's going to be a little different because it's the debut of Mikah Sargent doing it with me this year. So that'll be kind of fun. And actually I am kind of - I have a dog in this hunt because I lost my iPhone. Lisa cracked hers.

Steve: Whoa.

Leo: Well, I didn't lose it. I gave it to Lisa. Last week she cracked hers. And I said, I'm not buying you a new iPhone a week before an Apple event. That would be insane.

Steve: Right.

Leo: So I said, "Please, use mine." And, you know, I have enough phones to survive, I guess, a couple of weeks. But I'm iPhoneless for the duration.

Steve: You guess? Oh, my god.

Leo: I guess.

Steve: Yeah, uh-huh.

Leo: Should I use the Note 10 Plus or the Galaxy S10 Plus or the Pixel 3 or the...

Steve: You know, the only danger would be that expensing these might trigger a red flag at the IRS because they'd be saying, now, whoa, okay, now, wait a minute.

Leo: How many phones does a person need?

Steve: The first 50 phones we could allow. But, you know, when you are writing off the 51st phone, that's like, ah, yes.

Leo: This is a terrible time of year for me because I do have to buy all the new phones. We like to buy them instead of take loners because that way we get a real experience, you know, as a user.

Steve: Well, it's good that you're saying this because these tapes can be replayed to the IRS agent when they come.

Leo: Exactly. When the audit happens, you say...

Steve: And you say, look, look.

Leo: Please, watch Security Now! 731.

Steve: Had no choice. Now, that in no way relates to the fact that this podcast is titled "Deepfakes." Deepfakes has been in the news a lot, I mean, in the popular press because what's happening is, of course, with this crazy increase in technology that we're seeing, in available processing power - the concept of a photoshopped photo, I mean, "photoshopping" is now in the common vernacular because we know we cannot trust pictures any longer because they can be faked.

Well, what's happening now with the advent of all this computing technology is audio and video can be faked in real-time. Anyway, Facebook and Microsoft just announced an initiative to work on sort of counter AI that would be able to do deepfake detection. And so we've never talked about deepfakes before, and I thought that would be a great topic for us to wrap up the podcast.

But of course lots more is going on. We have the forced two-day recess of all schools throughout Flagstaff, Arizona last Thursday and Friday, which we have to talk about; the case of a ransomware operator being too greedy and what happened there; Apple's somewhat controversial response to Google's posting last week about the watering hole attacks that we talked about during last week's podcast. We have Zerodium's updated payout schedule and what it might mean for the industry. The final full public disclosure of BlueKeep exploitation code. There's no longer any mystery about how to execute remote code through this vulnerability that Microsoft released patches to, uncharacteristically, all the way back to Windows XP.

We also have some potentially serious flaws found and fixed, but not auto-updating because it really can't auto-update, in PHP that may require some of our listeners' attention. We've got a little bit of SQRL news, some miscellany, closing-the-loop feedback with a listener, and then you and I are going to talk about deepfakes. So I think another great podcast for our listeners.

Leo: Sounds like we have, you know, when I was a kid, the best thing that could happen to you was a snow day during the winter.

Steve: Yahoo!

Leo: It was a free day because of all the snow. I guess now we have to have ransom days.

Steve: Now we apparently do, yes.

Leo: School's closed. Wow.

Steve: And I have to say I kept looking back at their Facebook page all weekend and got incremental updates. In fact, that's our Picture of the Week is what was posted on Thursday, August 5th. So we'll be talking about that.

Leo: Wow. That's just crazy. Just crazy. Ransomware Monday.

Steve: So the news was posted to their Facebook page, probably on Wednesday. This is from the Flagstaff Unified School District. And it read: "Due to a cybersecurity issue that has impacted the ability of FUSD" - that's Flagstaff Unified School District schools in Flagstaff, Arizona - "to operate normally, there will be no school on Thursday, September 5th."

Leo: Woohoo.

Steve: And they said FACTS, which is probably, I don't know what, Flagstaff something or other [Family and Community Teaming for Students]. They said their childcare centers and FUSD, the Unified School District, preschools have also been canceled.

Leo: That bad.

Steve: Yes. And, not surprisingly, recess was then extended through Friday. They posted: "All Flagstaff Unified School District schools will be closed on Friday, September 6th, 2019 due to the continuing work to respond to the cybersecurity attack. Progress was made today in securing critical FUSD systems; but, unfortunately, work will need to continue through the weekend to ensure that students can return to school on Monday. The FACTS, childcare centers, and FUSD preschool remain closed on Friday, September 6th, as well. FUSD understands this decision impacts families and the community. We appreciate your patience as we work through this situation."

And so then I went, as I said, I was checking back through the weekend because I was curious how this was going. So they had a big "Important Message" banner that was up on their own page. They were cloning things over to Facebook also and just announcing that they were continuing to work. And they said, "An announcement regarding school on Monday [today], September 9th, will be made on Sunday, September 8th, when additional information is available." So of course they weren't and couldn't be sure until they were sure what the story would be.

And so toward the end of the day yesterday they updated their important message: "All FUSD schools will be back in session on Monday, September 9th. We appreciate your patience," blah blah blah. So at this point we don't have any additional information. But I did check this morning, and everybody's back in school. So the IT department for the Flagstaff Unified School District worked for some portion of four days in order to bring their systems back online. We don't know, again, as I said, anything. There's been no other information. Did they pay the ransom? Did they restore from backups? Was the impact minimal? You know, we have no more information.

But here, rather than talking about these events as we normally do after the fact, we watched it happen as this was going on. So yet another instance of a ransomware attack on a public entity having some consequence in the real world. There was, however, an interesting press conference held last Wednesday by Mayor John Mitchell, who's the mayor of New Bedford, Massachusetts. He held the first press conference of their adventure, which at that time was two months in the making, telling their interesting story of their own ransomware attack.

The city of New Bedford, Massachusetts' IT network was hit in the wee hours after the July 4th holiday, so like early, early on July 5th, with Ryuk ransomware, which by the way Malwarebytes now places at the top of their list of file-encrypting malware targeting businesses and municipalities. So it'll be interesting to see whether, moving forward, whether Sodinokibi's, as we've decided we should be pronouncing it, aka REvil, whether its affiliate marketing model is able to displace Ryuk as the number one threat in the ransomware business. But in any event, fortunately for the city of New Bedford, hackers breached that city's IT network and got Ryuk running on it in the wee hours of the morning following the U.S. annual Fourth of July holiday. Which was fortunate only inasmuch as apparently the IT staff came in early that day, maybe just taking advantage of the extended four-day weekend, to get some work done.

They discovered that the malware had at that point encrypted the files of 158 workstations, which, while significant, accounted for only 4% of the city's total fleet of 4,000 PCs. The attack would have been much worse had most of the city's systems not been offline at the time as a consequence of this extended four-day holiday. So that prevented the ransomware from spreading more deeply throughout the city's entire network. They were able to move quickly to disconnect the infected machines from the network so that they would not continue trying to infect more machines.

And in this press conference which was held last Wednesday, exactly two months following the attack, this Mayor John Mitchell said: "While the attack was still underway, the city, through its consultants, reached out to the attacker, which had provided by that time an email address." He said: "The attacker responded with a ransom demand specifically that it would provide a decryption key to unlock the encrypted files in return for a bitcoin payment equal to" - get ready for this number. $5.3 million they were hoping to get.

Now, as it happens, at the moment the city - apparently, it's not clear to me whether he said this or this was as the result of some reporting, but they didn't pay primarily because they didn't have that much money available. There was no $5.3 million with which to pay. If it had paid it, it would have made it the largest ransomware payment ever, which would have dwarfed the previous record, which was a million dollars paid by a South Korean web hosting firm not that long ago. I remember talking about them previously. But that didn't happen.

Knowing that they had no way of making that kind of a ransom payment, Mayor Mitchell said the city decided to engage in a conversation with the hackers primarily as a stalling tactic to give their IT staff more time to bolster the city's defenses. I'm sure that he was talking to IT, and they probably said, well, you know, we're really not sure how they got in, so give us a little time, if you can, to shore things up so that we can be sure they can't do more damage, like if they were actually more actively inside the network's system. So they stalled them a little bit, as much as they could, in order to keep them from running additional ransomware or whatever.

So anyway, at the press conference last Wednesday the mayor said: "In light of these considerations," he says, "I decided to make a counteroffer using our available insurance coverage in the amount of $400,000," he said, "which I determined to be consistent with other ransoms which had recently been paid by other municipalities." So of course I guess all city mayors at this point in the U.S. are probably aware of what the going rate is for ransomware. And clearly, 5.3 million was like, whoa, way over the top. We are seeing sometimes, we covered one a couple weeks ago, $600,000. And, you know, so 400,000 is in the ballpark.

Anyway, the attacker declined to negotiate at all, wouldn't even counter offer, and completely rejected the city's position outright. And since they wouldn't negotiate, and the city didn't have $5.3 million anyway, they decided to restore from backups, also since it was not a catastrophic 4,000 PCs that were hit, but only, as they said, 4% of their total fleet of PCs. So they restored from backup. It still took quite a while. I'm sure it was expensive. But it ended up being way less expensive than paying the ransom, which, as they said, they had no insurance coverage for anyway.

So here's a story of the people behind Ryuk, and we don't still know whether there are multiple attackers using Ryuk or one group somewhere who own Ryuk and are using it exactly in this fashion to attack municipalities. But it's probably served as a good lesson to the Ryuk people, the Ryuk operators that, well, you know, yeah, you can get multiple hundreds of thousands of dollars from a municipality, but not high millions. Not in the five million range. That's just not going to happen. So it turns out that the city probably did the right thing by saying no to these guys.

We do have a little bit of additional information on Texas, which as we know was hit by this Sodinokibi ransomware and which affected 22 municipalities. What we've learned is that the $2.5 million in ransom that was demanded by the Sodinokibi attackers was also declined, and that about half of the 22 municipalities that were attacked have managed to get themselves back online. Three weeks after the incident took place, the Texas Department of Information Resources (DIR) said that more than half of the impacted entities are now back to operations as usual. Some cities restored impacted systems from backups, while others rebuilt their networks from scratch. They just said, okay, I mean, you know, again, these were small municipalities who did have, we have verified, in common the use of a common cloud-based management backup facility. And so, as a consequence, they were able to avoid the ransom demand. So here again, $2.5 million? Nope, that's too much to ask.

The incident responders who managed the ransomware infections on behalf of these 22 Texas municipalities did publish some advice last week for companies and governments to follow. They had five points. They said: "Only allow authentication to remote access software from inside the provider's network." That, it turns out, was one of the problems. And actually one of our listeners had some detailed feedback about this particular instance which I've been unable to verify independently, and we'll share that when we get to it.

They also said: "Use two-factor authentication on remote admin tools and Virtual Private Network tunnels rather than remote desktop protocols." So that's interesting. You know, if we read between the lines, that suggests that maybe the problem was an RDP-based attack. But that's not what the feedback we have received suggests. So that's not clear, either. They also said: "Block inbound network traffic from Tor Exit Nodes. Block outbound network traffic to Pastebin. And use Endpoint Detection and Response (EDR) to detect PowerShell running unusual processes." So some good feedback from those guys.

And, you know, it's too early to say, but it feels a little bit like everybody now really has their guard up. It'll be interesting to see whether there was sort of a summer flurry of this which is going to slow down, or maybe whether the attacks move more toward corporate entities. And being nonpublic, if that happens, we may not be able to see that happening. We won't have the clarity that we have at the moment.

So last week we talked a lot about the interesting blog that Ian Beer at Google's Project Zero made where he told us about, basically, what they characterized as nondiscriminating waterhole attacks on a small number of servers targeting specific groups. And Leo, you had some information that had just been made public, I guess it was in...

Leo: TechCrunch had it.

Steve: TechCrunch, yes.

Leo: I think - Alex Stamos was tweeting about this. He suggested, and I think he's probably right, that Google gave this information on background to TechCrunch. Because Apple confirmed it.

Steve: Ah, okay, yes. And in fact we have some interesting code actually in the show notes that resulted from some additional research. I think it was RiskIQ that was able to provide additional detail. So anyway, so what we know is that last week Google published their blog about vulnerabilities that Apple fixed - oh, I'm sorry, what Apple posted at the end of last week in their Newsroom posting was, they said - oh, and I just got a big bright screen in my face. Windows 10 is back up.

Leo: Or just rebooting.

Steve: Yeah. No, it did that about five times; and then it gave me, like, we're updating stuff. Okay, so I'll turn that big bright screen off.

Anyway, so Google said last week - I'm sorry, Apple. Apple posted the following: "Last week, Google published a blog about vulnerabilities that Apple fixed for iOS users in February." They said, "We've heard from customers who were concerned by some of the claims, and we want to make sure all of our customers have the facts."

Apple wrote: "First, the sophisticated attack was narrowly focused, not a broad-based exploit of iPhones 'en masse' as described." So Apple was pushing back about some of what Google was saying. They wrote: "The attack affected fewer than a dozen websites" - and we'd heard 11 - "that focus on content related to" - and I can't pronounce this. What's the name of the Muslim community?

Leo: The Uighurs. The Uighurs.

Steve: The Uighurs.

Leo: Uighurs, yeah.

Steve: Uighurs community. They said: "Regardless of the scale of the attack, we take the safety and security of all users extremely seriously." They wrote: "Google's post, issued six months after iOS patches were released, creates the false impression of 'mass exploitation' to 'monitor the private activities of entire populations in real time,' stoking fear among iPhone users that their devices had been compromised. This was never the case," wrote Apple.

They said: "Second, all evidence indicates that these website attacks were only operational for a brief period, roughly two months, not two years as Google implies." They wrote: "We fixed the vulnerabilities in question in February, working extremely quickly" - which of course you and I, Leo, made a point of noting last week - "to resolve the issue just 10 days after we learned about it. When Google approached us, we were already in the process of fixing the exploited bugs." Which we didn't know of, if that's the case.

They said security - this is Apple - "is a never-ending journey, and our customers can be confident we are working for them. iOS security is unmatched because we take end-to-end responsibility for the security of our hardware and software. Our product security teams around the world are constantly iterating to introduce new protections and patch vulnerabilities as soon as they're found. We will never stop our tireless work to keep our users safe." So that was their blog posting. Which a lot of those in the tech community were a little put off by. They weren't really super happy with the tone of Apple's response.

Last week the head of Threat Research for RiskIQ told ZDNet for their reporting on this that the attacks were indeed very targeted, and that Google was wrong in its initial assessment. He later shared some thoughts in a public tweet, and provided us with some code.

Leo: Very targeted in their regard is, well, it could only have affected the one billion people who are in China.

Steve: Exactly.

Leo: Okay.

Steve: Exactly. So, yeah. Of course, as we know, that was the group. We have reason to believe that it was - and I think you had said this also last week, Leo, that it was China, no doubt the Chinese government that had infected some of these sites that was looking to pursue some of these groups. The JavaScript which was posted as part of RiskIQ's research actually shows them pulling the country of the user through JavaScript, looking for a match of the country equal to China. And if it's a Chinese person located at an IP address based in China, then it injects some additional JavaScript onto the page, all of this after waiting 60 seconds, which then performs this hack of iOS devices.

So, and actually I learned a little something from looking at this JavaScript code. There is a cool site, ip.nf, which if you visit, if you just put ip.nf into your browser, it pulls up a description of the API that this site provides. And if you put ip.nf/me.json, what you get back is a little bit of JSON, you know, JavaScript Object Notation, where some information about you, based on your IP, is shown. And when I did it, it was extremely accurate in its location.

And what the script does is it offers a little bit more parameterization. It's ip.nf/me.json?callback=jsons, which gives a more detailed result of sort of a full JSON hierarchy, and then the JavaScript parses that in order to get the country of origin of the user from which the user is connecting out of the JSON. And in my case it said "United States." And no doubt for people in China it comes back saying "China." So anyway, that's that story.

There was a very comprehensive analysis and reporting on Volexity, and I have the link in the show notes for anyone who's interested, who also noted that the same targeted attacks were also being waged against Android visitors.

Leo: Not the same ones. Different ones. Very different ones.

Steve: Well, and not overlapping.

Leo: Not at all.

Steve: So, right.

Leo: And we'd reported on that, I don't know if you did, but we talked about that. That was the attempt to put on everybody's phone who came into Xinjiang province, which is where the Uighurs are, East Turkistan, a malware, effectively a spyware application. So it's not the same, and I wouldn't expect Project Zero to even mention it. That's not what their mandate is.

Steve: Right.

Leo: I think Volexity kind of wanted to kind of somehow - I feel like it was an Apple apologist post. I didn't really think it was a right-on post.

Steve: Well, and of course it can't be the same attack because the two platforms are radically different.

Leo: No, it isn't. The only thing that's the same is they were targeting - it was apparently China targeting the same minority.

Steve: Right.

Leo: China's also throwing those people into concentration camps. Is that the same attack? No. It's not. And that's not what Project Zero is about. It's talking about zero-day exploits. So the question I want to ask you is, this word "indiscriminate" attack. Apple wanted it to - really wanted everybody to know, look, you're not a target of this attack. It's only people in China who aren't white. It's only Muslim people in China. Don't worry.

Steve: It's only targeted people. Thus it is clearly a targeted attack.

Leo: But if you target [crosstalk] as one billion people, is that, I mean, how targeted is that?

Steve: Good point, yes. I mean, we do know from this code that it would launch on your iPhone if you were in China. Period. Not a subregion. This code demonstrates very clearly that it was the country of China.

Leo: Right. So what in your estimation - so normally when I think of a targeted attack, I think, you know, China's found a dissident they want to - or some country's found a dissident they want to get into their stuff.

Steve: Good point.

Leo: That's a targeted attack. Or maybe a company that they want to get into. That's a targeted attack. It was indiscriminate in the sense that anybody in China who visited those 11 sites would be hacked, right, on an iPhone.

Steve: Yes, yes. And also remember that one of the reasons attacks are targeted is specifically to keep them from being discovered.

Leo: Right.

Steve: The likelihood of seeing an attack that was truly indiscriminate would be far higher than if it only affected people in China. So essentially it was a proscribed attack. Maybe that's...

Leo: There's a better word, "proscribed."

Steve: Yes.

Leo: It was not a universal attack. And it was targeted in the other sense that we often think of, you know, there are ransomware attacks that are aimed at anybody who happens to get the malware. So we mail it to everybody.

Steve: Right.

Leo: Those are clearly indiscriminate. And then there are ransomware attacks that are trying to get a specific company. Those are clearly targeted. In that sense, I guess it was targeted in the sense that they were trying to get certain people.

Steve: Right, right.

Leo: So the language doesn't live up to the description.

Steve: Well, yes. And I guess also there isn't really - this is an instance where we don't really have well-defined language for specifying the degree of targeting. It's certainly not an individual, and they were not targeting that population because they don't know how to do that. So they were just saying, well, at least we know they're going to be in China somewhere.

Leo: Right. And they're reading these Uighur news sites, or they're on a Uighur social network. So it's targeted to that respect.

Steve: Right, right. Anyway, so I guess we think, what? First of all, I completely agree with you that Project Zero is helping Apple. I mean, after all, we know they have reported privately more than 200 vulnerabilities to Apple. I mean, that's all for the good. That's helping everyone who is an iOS user, the fact that Google is going out of their way to improve the iOS platform. It's not as if they're saying, oh, you're on your own, Apple. We're going to only fix our own stuff.

Leo: Volexity and some implied that Google was doing this as anti-competitive FUD, like, oh, see, Apple's got problems; and Volexity said, but they ignored Google's problems. I don't think that's what Project Zero was doing. I don't think that's a fair characterization.

Steve: Yeah, I agree.

Leo: Yeah. And I think Apple's blog post was kind of unnecessarily critical of Google.

Steve: And they did not apologize. I mean, so...

Leo: Yeah, not really taking responsibility for this. And almost implying, well, don't worry, it was just these Muslims, no big deal. It wasn't a good thing.

Steve: Yes. Dan Goodin concluded his coverage for Ars by saying: "Another key criticism is that Apple's statement has the potential to alienate Project Zero, which according to a Google spokesman has to date privately reported more than 200 vulnerabilities to Apple." He wrote: "It's easy to imagine that it wasn't easy for Apple to read last week's deep-dive report publicly documenting what is easily the worst iOS security event in its 12-year history."

Leo: They would far prefer that Google had not said anything at all; right, yeah.

Steve: Yes. He said: "But publicly challenging a key ally on such minor details with no new evidence does not create the best optics for Apple." Which is really the point you were making, Leo.

Leo: Alex Stamos, who used to be the security guy at Yahoo! and Facebook, has said: "I've had a lot of experience with mishandled security flaws and long delays before revealing them." I think he got it right. He said that the real issue is the security teams at Apple and the security teams at Google, those are colleagues and have respect for each other. This was a PR release from marketing or legal. And he said: "This is not a path you want to take. Apple does some incredible security work, but this kind of legal and comms-driven response can undermine that work." So actually it was a message to Apple employees.

He says: "I've worked for companies that took too long to publicly address their responsibilities. Demand better." And I think that that actually is a reasonable point of view. This is a marcom response. Apple clearly didn't want the world to worry that their phones were insecure.

Steve: Right.

Leo: And so they thought that that maybe was the takeaway from Google's post. But the security professionals at both companies understand that's not really what's going on.

Steve: Yes. And when I looked at what Ian posted, at the technical details, I mean, there's no politics. There's no loyalty at the technology level. It was just - it's just simple facts. And I did feel like he was just saying this is what I found. And given the nature of this mistake in the code, I don't know how this could have gotten through QA. So he was just saying, well, I see what I see, and this is what I see. Without really putting a...

Leo: Project Zero skewers Google just as much. If they find a Google zero-day, they'll do the same thing. I don't think they're biased.

Steve: Yes. And remember that they were originally doing that. And it was when they opened themselves up to non-Google stuff that we all said, wow, holy crap, really.

Leo: These guys are good.

Steve: That's going to be really neat. And everybody has gained from it.

Leo: Yeah, I don't think it's any anticompetitive move at all on Project Zero's part.

Steve: Yeah. Zerodium, the controversial company we talk about from time to time...

Leo: Are they the good guys are the bad guys? I can never remember.

Steve: We're not sure.

Leo: Okay.

Steve: Yeah, we're not sure. It's funny because they have this chart of payouts which is meant to look like a chart of the elements. And then I realized, oh, yeah, of course, because Zerodium.

Leo: Oh, it's an element. Ah.

Steve: It's meant to sort of - it's supposed to be sort of an element, yeah. And the chart is not clear. You have to sort of study it for a while. But there is - they announced on September 3rd, last Tuesday, some major changes to their payout schedule. I first encountered this from a story at the Hacker News. And the way they put it I take a little bit of issue with. They said: "There's some good news for hackers and vulnerability hunters, though terrible news for Google, Android device manufacturers, and their billions of users worldwide." Okay, and that's what I sort of take issue with because I don't think it's ever bad to find bugs because they're going to get fixed if you find them. And the chance of them being found is greater if they're being exploited, even though, yes, there are victims of the exploitation until they're found.

Anyway, it appears that the zero-day marketplace has recently shifted toward the Android operating system with Zerodium now suddenly bumping payouts for researchers of the most severe class of Android zero-days up by a factor of 12.5. They were previously offering $200,000. They are now offering $2.5 million. Okay, hackers, take note: $2.5 million. So that moves it above the maximum $2 million payout for the equivalent gold standard exploit of that kind of bug for iOS. The gold standard is zero-click, persistent, root-level kernel access of a device, meaning that the user doesn't even have to click, doesn't even have to acknowledge or make a single click. A zero-click exploit is the gold standard for one of these.

So, and the reason I guess I take a little bit of issue with that characterization of calling it terrible news for Google, Android device manufacturers, and their billions of users worldwide is, you know, yes, if you were a member of an oppressed group or minority which some repressive regime might have a strong interest in monitoring, while also having that much money to spend in order to purchase what is arguably going to be a transient ability to do that, to monitor you, to intercept your communications or whatever, then yeah, you know, all other things being equal, the increased bounty on the Android side, and it is a significant increase, would likely increase the likelihood that hackers would be trying to pry into Android rather than iOS.

And the Hacker News notes, and I agree with them, they said: "Just like other traditional markets, the zero-day market is a game of supply, demand, and strategy, which suggests either the demand for Android zero-days has significantly increased, or somehow the Android OS is getting tougher to hack remotely," they said, "which seems unlikely." And of course, as we know, Zerodium is a controversial enterprise which purchases zero-day exploits from hackers and then, as far as we know, almost certainly resells them - because otherwise how could they afford to pay $2.5 million to hackers for exploits - resells them to law enforcement agencies and nation-sponsored spies around the world. So the only way the economic model would work would be if they were functioning as a middleman, essentially. So it seems that that is probably what's going on.

I got confused by my notes here because I kind of went off. A potential exploit purchaser might have said to Zerodium - yeah, right. I was sort of drawing a little bit of theory about how it is that they could be offering $2.5 million. So the theory would be that a potential exploit purchaser, a nation-state, for example, would say to Zerodium, look, we have a need. We will pay $3 million for 90 days of exclusive access to a zero-click persistent exploit for a fully patched Android device. Whereupon Zerodium turns around and makes this offer, which they did last week, subtracting their middleman commission, saying to the hacker/cracker community, okay, we're upping the ante for this class of exploit to $2.5 million. Give it your best.

So as I said, the same class of exploit on an iOS device is $2 million. So that's where the market is. In addition to this big change in Android, they also announced some app-targeted payouts, half a million dollars for submitting new persistence exploits or techniques for iOS and payouts for WhatsApp and iMessage. So now those two messaging apps for the first time are receiving additional attention for mistakes, not at the OS level, but at the app level, because of course that's where the action is also.

Leo: These prices are driven by customer demand, do you think? Or by difficulty? More by customer demand, I would think; right?

Steve: I would think yes. I would think by customer demand. Certainly they're not going to offer $2.5 million...

Leo: Unless they know they can sell it for that.

Steve: Exactly.

Leo: Yeah, yeah.

Steve: Exactly.

Leo: Sell it for twice that.

Steve: Either that, or they know that they have a demand pool that would pay, for example, a total of five.

Leo: Right.

Steve: So maybe they have five customers who they know they can sell that kind of exploit to for a million dollars each. In which case they know they're going to be able to double their money. I mean, it's clear they are a commercial entity. And so they've got an established customer base. They know what the going rate is for these things. And so, but something just happened. You know, something happened to move that from $200,000 on the Android platform to $2.5 million. And really, Android being open source, so hackers can look at it, it's much more difficult to attack iOS. It's a closed, to a much greater degree, a closed hardware and software ecosystem, so it made sense that it would be 10 times harder and therefore worth 10 times. On the other hand, a lot of people are using Android devices.

Leo: Yeah. I think I've also heard it said that...

Steve: The size of the market?

Leo: Yeah, well, often when you see this sudden jump, it's because some nation-state came along and said, you know, we really want to get those Uighurs, let's say. And most of them are using Android. We need an Android exploit, stat.

Steve: Yup.

Leo: And so it could also be that simple. There could actually be a buyer out there saying I need it right now. And boy, if you look at WhatsApp, of course. That's, I mean, that's what a nation-state would want to hack.

Steve: Yes. And Apple's maximum payout for this kind of exploit is $1 million. Okay, yeah. A lot of money. But Zerodium will pay two for the same thing.

Leo: By the way, that's also the point. They're going to always want to beat Apple.

Steve: Yes, yes. And so you've got to say, if you're a hacker who does manage, I mean, it's not easy to find an exploit now in iOS. Well, unless they make a bad mistake like they did recently with 12.4. But still, you know, would you rather double your money or just get a million dollars? It's like, no, I think I'd go for two. But it is when these things are found, I mean, and this is exactly the point of Google's Project Zero announcement last week that we covered, which was the result of several years of their study. These things are found when they are exploited in the wild.

So what's really interesting also is that it has to be the case that somebody paying $2 million, I mean, that has to be chump change, essentially, for the kind of purchaser of these things because they know by the act of exploiting this, there's a window before it gets discovered. And so they've got to be able to say, hey, we'll pay X amount of money, millions of dollars, for the opportunity for some period of time until it's seen and then reported and then patched, and then until the patch actually gets pushed out. Now, that's the other thing about Android, Leo. Consider how much more valuable an exploit is on Android devices that are not being patched because - so that also suggests...

Leo: Which is the majority of them.

Steve: Yes, yes. So that also suggests that there is much greater value to the holder of an unknown exploit. Even if it is patched, it's still going to hold its value because it only matters if your target is patching in order to close an exploit, not if Google is patching in order, you know, after they discover the problem. So that suggests there is a much longer tail on these things and that they have a lot more residual value over time.

So BlueKeep. We've been talking about this all year. And somewhat to everyone's bemusement, nothing has ever happened with this thing. It was, oh, my god, super critical problem. A readily exploitable flaw in the remote desktop protocol came to light at the beginning of the year. Microsoft was so freaked out they went all the way back to XP and offered patches for XP. This does not affect the newest versions of Windows, not 8.1 and 10, only 7 and Server 2012 R2 and previous. But still, it was trivial to cause a denial of service effect. It was much more difficult to leverage that into a remote code exploit.

But of course we're always talking about how the first thing that you discover is a way to crash the machine. And it's only then, after much greater analysis, and by applying much more finesse to the problem that you are, if you're sufficiently skilled, able to turn this into a remote code exploit.

Well, that happened last Friday. A Metasploit module was released on Friday, which puts it now in the open source community and public. So our friend Marcus Hutchins, aka MalwareTechBlog, posted a very clear expos. He titled his posting "BlueKeep," and this is the perfect way to frame it, too. He said: "A Journey from DoS to RCE." And he wrote, he started his posting: "Due to the serious risk of a BlueKeep-based worm, I've held back this write-up to avoid advancing the timeline. Now that a proof of concept for remote code execution has been released as part of Metasploit, I feel it's now safe for me to post this. This article will be a follow-on from my previous analysis."

So his previous analysis was posted back in May, simply titled "Analysis of CVE-2019-0708," which is BlueKeep. And that posting, his posting several months ago, begins, he says: "I held back this write-up until a proof of concept was publicly available, as not to cause any harm. Now that there are multiple denial-of-service proof of concepts on GitHub, I'm posting my analysis."

And so that was the so-called "denial of service" means you're crashing the remote server because other people who want to use it are denied its service because it crashed. So what we've seen is, although this took quite a while, we had the, okay, we now know how to remotely crash a server whose unpatched remote desktop protocol service is exposed. Then, as of Friday, now in the public, is the remote code execution. So this follows exactly the process that we've talked about.

I'm not going to go in any more depth because there's, again, it's very much like Ian Beer's analysis of the iOS stuff. It immediately drops you into the weeds. I got a kick, though, out of the fact that his intro to his second posting, which he just posted, picks up, like, right from where he left off. It's as if it's really one long posting, and he stopped with the first one exactly where what was known to the public stopped, where he went - he got to parity, and then the second one picks up, like with the next sentence.

So anyway, the takeaway for us is that the world now has access to everything needed for less skilled attackers to use the RDP protocol on still-unused machines to execute code on them. And of course this is much more worrisome now because I'm sure it's probably not even a pyramid. That is, if you were to graph the degree of expertise versus population, it's very rarefied at that high end, which is why it's taken until now for anyone to successfully reverse engineer a remote code execution. Remember that it was known it could be done, but they were very rarefied. And, boy, if you want to get a sense for how difficult it is, Marcus's write-up, it's a beautiful write-up that I have a link to in the show notes. So I recommend to anyone who's interested to take a look at it.

PHP has been moving along. As we know, this so-called Personal Home Page continues to be, by a huge margin, the number one most popular server-side web programming language in the world. It is able to lay claim to more than 78% of the Internet's servers. So it's significant when the maintainers of PHP release updates to it which patch multiple high-severity vulnerabilities in PHP's core and in some of its bundled libraries, noting that the most severe of these could allow remote code execution by attackers targeting and compromising servers.

So I wanted to give our listeners a heads-up. There are no publicly exposed services that PHP produces that automatically expose, like, all these servers in the world. It's not like RDP where there's a well-known service exposed. PHP operates behind the scenes as an enabling language in which web apps are written. But it does mean that, if the web apps your server is using uses some of these libraries which provide an exposure, then they're subject to arbitrary remote code execution. Quoting from the tech press about this, depending on the type, occurrence, and usage of the affected code base in a PHP application, successful exploitation of some of the most severe vulnerabilities could allow an attacker to execute arbitrary code in the context of the affected application with associated privileges. The most likely exploits or consequences are going to be denial of service conditions where you crash the system.

The tech press wrote: "The vulnerabilities could leave hundreds of thousands of web applications that rely on PHP open to code execution attacks, including websites powered by some of the most popular content management systems - WordPress, Drupal, TYPO3 and so forth." There are use-after-free code execution vulnerabilities in a regular expression library that comes bundled with PHP and is used by a number of programming languages.

So anyway, PHP doesn't update itself. My own server, I have a server at GRC where I run PHP. It's the core component, obviously, behind my own WordPress blog; the SQRL forums, which are XenForo forums, XenForo being a large PHP application. The little link shortener that I talked about recently is written in PHP, as are some other services. I've checked. My stuff is clean. It doesn't have - it's not using any of these vulnerabilities. But I nevertheless updated because I would recommend everyone to do that.

It's not something that can happen automatically because there isn't an automatic update service for PHP. The process, obviously, for anyone who's maintaining PHP, I'm sure they know, you have to bring the server down, then update the core PHP components because the server is using those, often keeping them in use in memory, depending upon how you are causing PHP to be executed. And then bringing the server back up after you've got PHP updated. So anyway, my point is it's not a super emergency, despite the fact that it is, like, incredibly widely used on the Internet.

Probably, if you're the party responsible for maintaining PHP, you're aware of how to update it. I just did want to put it on everyone's radar that there was just recently a round of critical updates affecting the 7.3 chain, the 7.2 chain, and the 7.1 chain. So when you can, it's worth doing.

Since last week I wanted to mention that the SQRL documentation project is completed. Four documents which are 79 pages across four individual PDFs are now in place, which completely explain SQRL's features, its operation, and its implementation in sufficient detail for anyone interested to create SQRL clients and servers. There's still some discussion over in the SQRL newsgroup about some little edge conditions around some condition flags. So as soon as I'm through with the podcast, I'm going to get back to get that final stuff squared away. But basically the documentation project is finished.

Also Firefox, I'm noticing, is getting some increasing attention. I know you, Leo, have been looking at it more closely, liking it, after you've gotten sort of used to the way it differs from Chrome. I've heard you talking on some of your other podcasts and to different groups about Firefox. I have a machine which is RAM constrained, which is funny to say about a machine with 8GB these days.

Leo: That's funny, yeah.

Steve: Isn't that unbelievable?

Leo: Yeah, yeah. But true, it's true.

Steve: Oh, my goodness. It is. And so sometimes when I'm preparing the podcast on that, it's an older Lenovo X1 Carbon. And I could only get 8GB. That's the most...

Leo: I remember when you bought that, yeah, yeah.

Steve: Yup. And I still love that machine.

Leo: No, it's a great machine, yeah.

Steve: I know that the X1 Carbon is your favorite machine, too. If I were to buy another one, and I eventually will, I again will get the maximum memory, which now is 16GB. And I probably wouldn't have a problem if I were to have that. But I'm sometimes getting warnings that my system is out of RAM when I have many tabs open in Firefox. And - I know.

Leo: You must have so many tabs.

Steve: Well, actually, they are - I think they're large pages. And so, for example, BleepingComputer has very large, very content-heavy pages. And so what I'm doing is I'm going through stories, clicking on pages that I will then come back and read. So my work methodology has me open a bunch of pages, which I'm then going back to examine and read and study and decide if they make it onto the podcast. All of this by way of saying that I was reminded that Firefox, in I think it was 67 - we're now at 69. And 67 was supposed to have obsoleted the manual release page, release tab memory feature. But it apparently isn't working. Or maybe it isn't fast enough or aggressive enough. Or it doesn't look to see whether the system is running out of memory. I don't know what the problem is. But I'm still running out of memory.

And in the show notes, Leo, I have a screenshot of Task Manager running in Windows when I was nearly out of memory and I did something in Firefox. And so this is a Firefox user tip. There is a very lightweight add-on called "UnloadTabs." And the option it offers is unload all but the current tab. And so you right-click on the current tab, down on the bottom. Unload all but this one. You click it, and you can see what happens to memory. I mean, it almost dropped, well, not quite in half. I've seen it drop in half. It makes an immediate release; and then a few seconds later, I don't know, some other stuff gets released, and then it makes for a nice savings.

So anyway, I just wanted to pass this as a tip to our listeners. There are a couple of them. UnloadTabs was recently updated to reflect the change that Firefox made to their add-on system so it again - it stopped working. The author invested 40 hours of time bringing it back. And I'm thankful that he did because it's handy just to be able to say, if you have a bunch of tabs open, they can be using a lot of memory. And so this just allows them to just - the tabs stay, so as placeholders, which is all I'm using them for. You get to keep your placeholders. But you'll only then reload the memory as you then click the tab to bring it current again. So just a little tip for our listeners.

Speaking of tips, I found - it was a tweet from a Notre Poubelle. On Saturday he tweeted: "Hi, Steve. I have a SpinRite question. I am not currently a SpinRite customer." He says: "I have a hard drive that I suspect is having problems. I ran Windows chkdsk C: /f /r /x a couple of times, and it gets stuck for a very long time at the same percentage point. In this case, I actually don't care at all about the content of the hard drive." Oh, that's interesting. So it's C:, but he still doesn't care about it. That's cool. He says: "My question is, is it worth purchasing SpinRite, given that I don't care about the contents of the hard drive? Or would it be wiser to just buy another hard drive? Assuming I did buy SpinRite, and it fixed the hard drive, is it likely a temporary fix, and I'll end up having to buy a hard drive anyway?" Great question.

Leo: That is. That is. Because hard drives are so cheap now.

Steve: Yes, yes. And so that was one of the things I thought was that hard drives are cheap. On the other hand, you have to pay for a cheap hard drive every time you purchase a new one. You only have to pay for SpinRite once.

Leo: True. Good point.

Steve: And it will fix all the hard drives you have every time they get in trouble, until they finally do really die. And it is the case, I mean, first of all, there's no question that SpinRite would get past that sticky point, and it might be, for example - well, for example. Say that the drive was writing when the power failed on it, like if it was an external drive, and you pulled the cord. Well, that would definitely damage the sector that was being written at the time. But that by no means means that the hard drive is no longer any good. It just means that that spot, that sector, while being written, didn't get its checksum set correctly.

So SpinRite will fix that. It'll either fix the checksum - if writing it is able to fix it, SpinRite will do that. If rewriting it won't fix it, then SpinRite will get the drive to relocate it, basically removing it from service forever and replacing it with a good one. So your drive will get fixed. And SpinRite will do this every time something happens to any drive you own. So again, and of course in the case of SpinRite, as you know, then you'll be able to get all of the other data that was hiding behind that bad spot, especially, for example, if it happened to be a directory sector. Then you would lose access without SpinRite to everything else that was downstream of that point in your directory hierarchy.

But again, this guy said he didn't care what was on the drive. So it's certainly a bit of a tradeoff. For what it's worth, SpinRite would fix it, and the drive is probably okay. And the other cool thing is that, when you're running SpinRite, that SMART data will show you if the drive is getting soft. If SpinRite pushes the health parameters down, which exposes red bars in the SMART data's bar chart, that's a sign this drive is getting soft; and, yes, SpinRite is then advising you that it's probably good to replace the drive. So anyway, great question, and not a simple answer.

Leo: I like it. So it kind of depends why you're losing sectors. But I like it that SpinRite'll give you some idea of whether the drive is healthy or not.

Steve: Yeah, yeah. It really is great about doing that.

Leo: That's handy.

Steve: And that's the other thing people miss is that that SMART, the Self Monitoring Analysis and Reporting Technology, SMART, it's only useful if the drive is being worked, if it's under load. If it's being asked to do work, that's where you see sectors being relocated, or seeks having to be retried too much, which is an indication that the low-level format is getting soft. So you have to put it under load. And so SpinRite together with that SMART analysis is just an incredibly powerful tool.

Anyway, Steve S. tweeted me. He says: "I'm an avid listener of Security Now! and wanted you to know that you are correct about the 22 towns that were infected with ransomware. A third party was providing IT services" - now, that we knew last week - "access to state DMV resources, utility payments, et cetera." So he has some more information than we had. And he continues: "The third party had an OpenVPN connection to each municipality. Not sure where it started, but it spread everywhere. Keep up the good work. I enjoy the podcast."

And so that's interesting because we've only talked about, and we generally talk about, OpenVPN in the context of on-demand. I still want to use the term "dial-up," you know, because that's the way it was once upon a time. But, for example, that's the way I'm still using or I'm currently using OpenVPN. I normally don't maintain a static connection bridging networks together because I don't have a need. I will bring up an OpenVPN tunnel the way - and that's the way, of course, most people use OpenVPN, when you're out roaming, and you want to connect to your home base, to your corporation, to their OpenVPN services. You bring up an OpenVPN tunnel. It establishes. You authenticate. And then you're now part of that network on the other end of the tunnel.

But the other way to use OpenVPN, and it is arguably, I mean, it's robust, and it's a cool approach for those who need it, is for it to be used in a static network bridging mode where - and again, it does this really well, where you have two endpoints, and they establish a link across the Internet, bridging their different subnets, basically their non-Internet LANs into a single LAN. And OpenVPN will even transmit broadcasts. So you can create a common broadcast domain across disparate LANs so that you can do full discovery of systems and services across that link. It is very cool. So it doesn't just run at level 3. It's able to run at level 2 and actually create an Ethernet bridge, not an IP bridge, between separate networks.

So anyway, in this instance, it looks like, from what Steve S. has tweeted, that what this group probably had was a bunch of static OpenVPN connections from this IT service out to these 22 municipalities. And so we don't know where the - we don't know that the IT service even themselves was the source of the infection. It may have been one of those 22 municipalities, in fact, it seems more likely, one of those 22 got infected. The malware rode up the VPN connection into the IT service and then back out. It went, aha! It's like, it struck gold, then ran out the other 21 VPN connections to get into all of the other 21 municipalities and brought them all down. So a real interesting case of an OpenVPN-enabled network dramatically magnifying the size of an attack.

But even so, the bad guys didn't get paid. They asked for 2.5 million. These guys, the small municipalities said, no thanks, we're not that big. The damage done wasn't that extreme. So we'll restore from backup where we can, or we'll just rebuild our networks. Thanks very much.

Leo: Happy ending. Sort of.

Steve: Happy ending.

Leo: Sort of.

Steve: Oh, yeah, well, yeah.

Leo: Now let's talk about Deepfakes.

Steve: Deepfakes. So this first, well, I've been aware of it happening for a while. I've been noticing that there's been a lot of conversation about it and demos and things on standard broadcast TV. A story caught my attention that I was going to cover this week. And I decided when I saw the second story about a move being made to create deepfake detection, that it was worth just pushing back a little bit and spending some time talking about it. The first story was in The Wall Street Journal, which is behind a paywall. So I was able to get the first paragraph of it and then dug around and found some more information from others who have Wall Street Journal subscriptions. The Wall Street Journal story was titled "Fraudsters Use AI to Mimic CEO's Voice in Unusual Cybercrime Case."

Leo: This was wild; wasn't it? Oh, my god.

Steve: Yes. And their coverage said: "Criminals used artificial intelligence-based software to impersonate a chief executive's voice and demand a fraudulent transfer of 220,000 euros ($243,000) in March in what cybercrime experts described as an unusual case of artificial intelligence being used in hacking. The CEO of a U.K.-based energy firm thought he was speaking on the phone with his boss, the chief executive of the firm's German parent company, who asked him to send the funds to a Hungarian supplier. The caller said the request was urgent and directed the executive to make this transfer."

So this sort of put me in mind of the way the world is changing. In the very early days of email, the very, very early days, we might have taken a note from apparently our boss at face value and acted upon it. But obviously those days are long gone. And for some time, due to its power to convincingly fool and spoof, the term "photoshopping," as I mentioned at the top of the show, has become a well-established term now. Everyone knows what a "photoshopped image" means, and Adobe gets a lot of credit for Photoshop being the app behind that.

But now, as I said at the top of the show, the crazy power of today's computational resources has moved us to an entirely new level of spoofing. Not only are we able, we have been long-time, been able to spoof a static image, now we're able to spoof, and this is what has changed so much, complex time-varying signals, both audio and video, in real-time.

So today, upon receiving something important and unexpected via email, we might give our boss a call and say, hey, did you just send me an email asking, you know, whatever. But notice how we would confirm it. We would confirm by voice because today we still trust that. So the point here is no one trusts a photo anymore. That's been spoiled. And we're also on the brink of losing our ability to trust real-time non-face-to-face electronic communication, as well.

We're entering a future where it might be necessary for the instructions to be, if I ever tell you by email or even by phone, I guess even in a video conference, to do something you don't expect, I want you to confirm it face to face. You know, just walk down the hall, stick your head in the door, and ask, "Hey, did you just call me and ask me to something?" And I guess my point is, think, Leo, how much that changes our world, if we can't trust, not only our own ears, but our own eyes.

Leo: A number of online brokers, I know Charles Schwab does this, and I know Fidelity do this, have voice authentication. And I always thought that was a little sketch. But, you know, you still have to have the password. It's like a third factor.

Steve: So you mean, so they try to do voice recognition?

Leo: Yes, on you.

Steve: Wow.

Leo: You have to train it before you use it, obviously. So that's an option. And I've always just ignored it because I thought, well, my voice has been recorded so many times, I don't think it'd be too hard to synthesize, you know.

Steve: Well, and of course that's where the whole AI thing comes in; right? Because there's this notion of training and AI to be able to sound like Leo.

Leo: I have to say, though, the giveaway will be the content, not so much my voice. You could make something sound exactly like me. I've got John Legend coming out of my Google Assistant you know. That's one of the voices now you can use for your Google Assistant is John Legend. And it sounds just like him. But you can still tell, partly because of the content, partly because voice synthesis isn't perfect, that it's not actually John Legend.

Steve: Well, and good point. For example, some of the spam email you get, you look at, and it's like, okay.

Leo: Yeah. That isn't a human.

Steve: This is not my mom who sent this.

Leo: It's grammatically incorrect, you know.

Steve: Exactly.

Leo: So I think we still have a way to go, even those deepfakes. You've seen some of these deepfakes.

Steve: Oh, yeah.

Leo: They're good, but they're not perfect. You can tell.

Steve: Yes, yes.

Leo: I guess really what we have to prepare for is a day when they are perfect because photoshopping is completely, you know, it's very hard to detect.

Steve: That's a very good point. And I think we have to acknowledge, and that's why it's worth talking about, that that day is coming. And that we know there are entities that are willing to invest in that technology. I mean, certainly academically, it's an interesting academic exercise to see whether you can create that kind of fake. But, for example, the U.S. election, we know that Russia messed with our past presidential election in order to stir things up in various ways. And there's every indication that, hey, they were successful, at least in creating some chaos.

Leo: Thousands of Twitter bots. A Twitter bot is easy, is an easy thing to fake. And a Facebook bot. That's not hard.

Steve: Yes. So the second part of this is an interesting challenge that has just been announced. It's called the Deepfake Detection Challenge, DeepfakeDetectionChallenge.ai. I thought that was a very cool domain for it to have. All one word, or all run together: DeepfakeDetectionChallenge.ai. The subhead reads: "Deepfake Detection Challenge invites people around the world to build innovative new technologies that can help detect deepfakes" - good luck with that - "and tampered media. Identifying tampered content," they write, "is technically challenging as deepfakes rapidly evolve, so we're working together to build better detection tools." And of course this is going to be cat and mouse; right? Because all of this is going to be done in public. All of the detection stuff is going to be done in public. The test sets will be public. And so the people making the deepfakes will also be able to check their fakes against the detection systems.

Anyway, Facebook and Microsoft, in partnership with academics from Cornell Tech, MIT, University of Oxford, UC Berkeley, University of Maryland, College Park, and University of Albany, New York, have joined forces to sponsor a contest promoting research and development to combat deepfakes. Videos altered through - well, videos because audio, if you can do video, you can obviously also do audio - altered through artificial intelligence to mislead their viewers.

So we have the DFDC, as it's been called or being called, the Deepfake Detection Challenge, which aims to spur the industry to create technology that can detect and prevent deepfakes, according to a blog post by Facebook's CTO, Mike Schroepfer. Mike posted under a title "Creating a dataset and challenge for deepfakes." So this is the Facebook CTO who said: "Datasets and benchmarks have been some of the most effective tools to speed progress in AI. Our current renaissance in deep learning has been fueled in part by the ImageNet benchmark. Recent advances in natural language processing have been hastened by the GLUE and SuperGLUE benchmarks."

He says: "Deepfake techniques, which present realistic AI-generated videos of real people doing and saying fictional things, have significant" - and you can imagine why Facebook is particularly sensitive to this now.

Leo: Oh, yeah.

Steve: Uh-huh, "...have significant implications for determining the legitimacy of information presented online." And of course, Leo, this is why it occurred to me that this also is a little relevant to the discussion we're going to be having in Boston on October 3rd.

Leo: Oh, yeah.

Steve: Because it's about identity. And of course this is spoofing identity all the way up to a video that is convincing. So he says: "Yet the industry doesn't have a great dataset or benchmark for detecting these fakes. We want to catalyze more research and development in this area and ensure that there are better open source tools to detect deepfakes. That's why Facebook, the Partnership on AI, Microsoft, and academics from" - and all those universities that I just mentioned - "are coming together to build the Deepfake Detection Challenge."

He says: "The goal of the challenge is to provide technology that everyone can use to better detect when AI has been used to alter" - god, this will be really interesting - "to alter a video in order to mislead the viewer. The Deepfake Detection Challenge will include a dataset and leaderboard, as well as grants and awards, to spur the industry" - and I saw elsewhere, oh, yeah, it's later in his posting - "to spur the industry to create new ways of detecting and preventing media manipulated via AI from being used to mislead others. The governance of the challenge will be facilitated and overseen by the Partnership on AI's new Steering Committee on AI and Media Integrity, which is made up of a broad cross-sector coalition of organizations including Facebook, WITNESS, Microsoft, and others in civil society and the technology, media, and academic communities.

"It's important to have data that is freely available for the community to use, with clearly consenting participants and few restrictions on usage. That's why Facebook is commissioning a realistic dataset that will use paid actors with the required consent obtained to contribute to the challenge. No Facebook user data will be used in this dataset. We are also funding research collaborations and prizes for the challenge to help encourage more participation. In total, we are dedicating more than $10 million to fund this industry-wide effort. To ensure the quality of the dataset and challenge parameters, they will initially be tested through a targeted technical working session this October at the International Conference on Computer Vision.

"The full dataset release and the DFDC launch will happen at the Conference on Neural Information Processing Systems (NeurIPS) this December. Facebook will also enter the challenge, but not accept any financial prize. Follow our website for regular updates. This is a constant evolving problem," he finishes, "much like spam or other adversarial challenges. And our hope is that by helping the industry and AI community come together, we can make faster progress." So, wow.

Leo: Interesting.

Steve: Yeah. It is. This is very cool. It's clear that this is going to engage the interest of the industry's top AI academics and dynamic media people to see whether we can come up with some means for detecting this. And of course it's also going to have an effect of providing feedback to the bad guys, who look at how this detection is being done and then see if they're not able to avoid being detected.

Leo: You know, it strikes me that the human brain is marvelously attuned to detect fakes, to detect things that seem off. And it's an evolutionary, I think an evolutionary ability because, if people are ill, you want to stay well away. It's a survival instinct. And so anything slightly off, that's the "uncanny valley," right, when we do animation, and we make it look very human. We can tell it's not human. There's just something subtle that maybe we're better at than any machine will be. But what I was thinking is the real issue with deepfakes is not a three-minute video of the President saying we're going to attack Iran. The real deepfake is a subtle, maybe a one-word change to an existing video.

Steve: Yes.

Leo: That would be - you would be very hard pressed to detect that if it's done right. I think we have the tools to do it now.

Steve: Yes. Very, very, very, very good point, where you just make a...

Leo: Just a little.

Steve: A subtle alteration that has dramatic magnifying, sort of multiplicative effect.

Leo: Remember the video that the administration proposed when they banned Jim Acosta, the CNN reporter from the White House press briefings, which now of course they don't do anymore, so it doesn't matter, but they had a video of him pushing an aide away. She was trying to take the microphone. And it was a modified video of him pushing the aide away. People figured out it was modified. But it was such a subtle little thing. And that's the key is, if you just make a little difference, no, that puck made it into the net. See, right there. That would be a lot harder, I think, to detect than a longer video.

Steve: Well, yes. And even from an automated standpoint, so I think your point is not only for humans to detect, but for, like, I mean, AI is going to have to look at the whole thing, frame by frame.

Leo: Yeah, what am I looking at, yeah, right.

Steve: Yeah. Yeah, I would argue that - and I think you're right, Leo. Humans, due to our nature, I mean, you know, for example, think about how like in poker there are tells.

Leo: Oh, yeah. Oh, yeah.

Steve: And you're able, like a good poker player is able to look at something that isn't fake, but it's just a tiny little tick that indicates something going on inside the person's head. I mean, wow.

Leo: We're very good. That's our whole biology is tuned to do that. And it's a little harder for a computer, which can't even really - a two year old can recognize his mother's face. A computer has to work a lot harder. It's a good challenge, though.

Steve: Well, and how many spouses are sure they can tell when their spouse is not telling them quite the whole truth.

Leo: Yeah. That's exactly right. That's exactly right. We've never made a good lie detector. There's never yet been a lie detector that actually works. Better than a spouse, anyway. Great subject. I really look forward to - see, we're going to be doing this for a long time now. Couple hundred more episodes.

Steve: That's right.

Leo: By then we'll have lots more to talk about with deepfakes and everything else. Poor Steve. You're really getting to the end of the rope, aren't you. You're starting to look ahead to 999.

Steve: I have a feeling maybe we're going to have to do some renumbering here somehow.

Leo: Really? Are you getting tired?

Steve: Ah. What? No. No.

Leo: No, come on. You're going to have a rebellion on your hands if you say I'm quitting or retiring. I do want to mention, you mentioned our October 3rd event that's coming up. That's very exciting. Steve will be relaxed because he'll just be back from a few weeks in Europe, right, as you're touring. The SQRL Tour heads to Dublin and Copenhagen?

Steve: Yup, Dublin and then Gothenburg, Sweden.

Leo: Goteborg in Sweden, yeah.

Steve: Yup, yup.

Leo: So we'll be in Boston October 3rd, 3:45 p.m. It's a summit. I think it'll be a great panel discussion about this whole problem of authentication, how you prove who you are, including not just Steve. I'll be there, the CISO Gerald Beuchelt from LogMeIn will be there, and the legendary Bill Cheswick (Ches) will be there. And, I mean, we've got four people who have spent a lot of time thinking about what's next after passwords. If you would like to go, it's free. You do have to obviously be in Boston. And if you're not going to be in Boston, never fret. We're going to make a video and post it online. It'll be a TWiT Special.

But if you want to get tickets, if you are going to be in Boston on October 3rd, go to twit.to, that's our URL shortener. Not TWiT.tv, twit.to/unlocked, twit.to/unlocked. You could fill out the form there. I think there's still room. We've had to expand the auditorium. And LastPass, who's our sponsor, says they're willing to expand it to the point that they get too big for the Intercontinental Hotel, and at that point we'll have to stop. But there are bigger ballrooms, and we're getting them.

What's really cool is it's a benefit. It's a charity. LastPass is donating $100 on behalf of each attendee to the Boys & Girls Club of Boston, the Greater Boston Food Bank, or KodeConnect, and you get to choose. So this is a really neat event, and it's going to be a lot of fun. There's a cocktail party afterwards. I can't promise Steve will stick around, but maybe we can persuade him to.

Steve: Oh, I will for sure.

Leo: If it helps, I'll bring a cardboard cutout of Captain Kirk and - too soon?

Steve: It'll always be too soon.

Leo: Find out more.

Steve: I'd had nothing to drink, Leo.

Leo: No, that's the funny thing. He doesn't need it. Twit.to/unlocked. That's the short URL. And there's still room; so, please, we'd love to see you on October 3rd in Boston.

This is an unusual record time because of tomorrow's Apple event. Thank you, Steve, for being flexible. We'll be back on Tuesdays, 1:30 Pacific - what?

Steve: We also have another unusual record time on Saturday.

Leo: Oh, I forgot, yeah. Because, again, this European tour, the grand tour that Steve's doing, we're going to record two episodes ahead of time. This Saturday; right?

Steve: Yes. It will be "The Joy of Sync."

Leo: We're going to find out what Steve finally came up with in his quest for the perfect file sync solution. I think this is - I can't wait for that. So that's Security Now!. That'll be a pre-record for the following week. So, yeah, we're going to be all messed up, aren't we.

Steve: Yeah.

Leo: So this Tuesday we're doing a show, 1:30 Pacific, 4:30 Eastern, 17:30 UTC. Not this Tuesday, a week from tomorrow.

Steve: Correct.

Leo: So confusing. So September 17th.

Steve: Because we've got Apple Day tomorrow.

Leo: Yeah, Apple Day tomorrow. September 17th we'll be back on schedule. But September 21st we'll be recording for September 24th, that's a Saturday afternoon right after the radio show. And then you'll be gone for a couple of weeks. Is that right?

Steve: Yes. Correct.

Leo: All right. All right.

Steve: And then Episode 734 is where we will air the "Joy of Sync" episode that we record this first Saturday.

Leo: This Saturday, yeah.

Steve: Yes.

Leo: So I'll be spending a lot of time with you over the next two weeks.

Steve: That's going to be great.

Leo: September 14th, this Saturday, we'll do another one. That'll be for 10/1. That'll be for October 1st. And that'll be a lot of fun.

Steve: Exactly. Exactly.

Leo: I'm sure I've just confused the heck out of everybody. We'll be here a week from Tuesday and for the next two Saturdays.

Steve: And if they subscribe to the podcast...

Leo: That's true. Don't have to worry.

Steve: Never fear, they'll get them all.

Leo: It'll all be the same, just click the button at TWiT.tv/sn, subscribe to the podcast. Steve, of course, has copies of every show, 16Kb for the bandwidth impaired, 64Kb for those of you who like the full sound. And of course very handy, the transcripts that Elaine Farris does for each and every episode. Those are all at GRC.com. While you're there, pick up a copy of SpinRite, Steve's bread and butter and the world's best hard drive maintenance and recovery utility. There's lots of great stuff, too. Just, you know, if you're going to GRC.com, block out some time. Get some tabs ready because there's a lot of great stuff there. You might want to turn on that UnloadTabs out-of-memory thing before you do that because there's so much good stuff there.

We have it on our site, as I mentioned, TWiT.tv/sn. And if you subscribe, then you don't have to worry about anything. It'll just automatically come to you whenever it's done. Thank you, Steve. Have a great week. I'll see you a week from tomorrow.

Steve: Right-o.

Leo: I'll see you Saturday.

Steve: Actually you will, yes. I was just going to go with it because you'll also see me a week from tomorrow.

Leo: It's so confusing. I'll see you Saturday.

Steve: Okay, buddy.

Leo: Take care, Steve.