EPISODE 779

[INTRODUCTION]

[00:00:00] JM: Advertising fraud steals billions of dollars every year. BuzzFeed reporter, Craig Silverman, reports on advertising fraud and its impact on the internet. In one investigation, Craig uncovered a mobile advertising fraud scheme in which four people stole millions of dollars, perhaps as much as 75 million, or even 750 million. We don't know the exact numbers, but they did so by serving advertisements to automated users on mobile applications. The scheme worked as follows; a shell company called We Purchase Apps would buy legitimate applications from legitimate app developers. Then the new owners of this legitimate application would record the behavior of the real users on those apps.

So imagine a trivia game where somebody has developed a trivia game, some people are downloading it, some people are using it and the trivia game has monetized with ads. This company, We Purchase Apps, would buy this real legitimate app from those creators who have made the trivia app and then they use the behavior of the users on those apps to train models of fake users who could replicate that behavior.

So imagine somebody purchases this trivia app from the original creators of the trivia app and then they record the user behavior and then they train machine learning models based off of that legitimate behavior. Now that you've got a user model, you can just spin up as many fake users as you want and those fake users are going to consume advertisements just like real users. Then the new owners of this trivia app would earn all the money generated by displaying ads in these apps. This is a very simple scheme. It's easy to pull off. It did not require much sophistication in terms of engineering or business skills, and it took place across many, many mobile applications. If a group of four people can generate tens of millions of dollars, how much ill-gotten capital is being generated by large corporations that are deeply involved in the advertising market?

Craig's article went viral and he has followed it up with several other pieces about ad networks, fraud investigations by Google and the potential for mobile apps to be used for large-scale surveillance of Americans by the Chinese. Craig is the most dedicated reporter covering

advertising fraud today. His work is invaluable, because he's asking difficult questions about the economics of our internet.

As we discuss in the episode, there is currently no effective means of automatically detecting a bot from a human on the internet. Consider the ramifications of this; we cannot detect who is a bot and who is a human. We've discussed this in detail on previous episodes about advertising fraud, the advertising industry as a whole, advertising analytics and the various techniques of ad fraud. Ad fraud is not the fault of any one party. There is not a big, bad villain that is orchestrating all of the add fraud. It's an emergent result of the way that our internet is set up, and it's as hard to imagine a world without advertising fraud as it is to imagine a world without email spam.

I really enjoyed talking to Craig, and I hope you enjoy the episode as well.

[SPONSOR MESSAGE]

[00:03:45] JM: Triplebyte fast-tracks your path to a great new career. Take the Triplebyte quiz and interview and then skip straight to final interview opportunities with over 450 top tech companies, such as Dropbox, Asana and Reddit. After you're in the Triplebyte system, you stay there, saving you tons of time and energy.

We ran an experiment earlier this year and Software Engineering Daily listeners who have taken the test are three times more likely to be in their top bracket of quiz scores. So take the quiz yourself anytime even just for fun at triplebyte.com/sedaily. It's free for engineers, and as you make it through the process, Triplebyte will even cover the cost of your flights and hotels for final interviews at the hiring companies. That's pretty sweet.

Triplebyte helps engineers identify high-growth opportunities, get a foot in the door and negotiate multiple offers. I recommend checking out triplebyte.com/sedaily, because going through the hiring process is really painful and really time-consuming. So Triplebyte saves you a lot of time. I'm a big fan of what they're doing over there and they're also doing a lot of research. You can check out the Triplebyte blog. You can check out some of the episodes we've done with Triplebyte founders. It's just a fascinating company and I think they're doing something that’s

really useful to engineers. So check out Triplebyte. That's T-R-I-P-L-E-B-Y-T-E.com/sedaily. Triplebyte. Byte as in 8 bits.

Thanks to Triplebyte, and check it out.

[INTERVIEW]

[00:05:34] JM: Craig Silverman, you are a journalist with BuzzFeed. Welcome to the show.

[00:05:38] CS: Thank you.

[00:05:40] JM: You investigated a large advertising fraud ring that stole tens of millions of dollars, and we'll get into the mechanisms of the ad fraud ring, but I want to start with why you found yourself researching advertising fraud. How did you get into that subject?

[00:05:57] CS: Right. So it started close to two years ago, and my background is that I have been looking at different types of digital deception and media manipulation for a while. This is sort of my beat and my focus. A lot of that has been around online misinformation. So completely false story spreading on Facebook and other platforms, trolling campaign, bots, all the different ways that this new media environment we have can be exploited and manipulated.

So as at the end of 2016, I was of course writing a lot about political misinformation and did a lot of recording there, not only in 2016, but long before that. Then as we got into 2017, I was sort of casting about and saying, "IF things are so messed up there, where else are they messed up?" and people started telling me, "You need to look at advertising. You need to look at ad fraud," and I had no idea really what was going on with that and quickly realized, "Oh, wait a second. People are stealing billions and billions of dollars and nobody is going to jail. People aren't losing their jobs. The industry has just seem to have come to accept it," and for me as a journalist, you look at that and you say, "Well, there must be tons of stories there and nobody's writing about them," because while some people of course in the industry do care, there's really been no scrutiny about it from the press, which just blew my mind.

[00:07:16] JM: And why do you think that is? Why hasn't there been scrutiny from the press?

[00:07:20] CS: Well, I think there's probably a couple of factors. So one is that it is a technical thing, and for a journalist to kind of get up to speed on it and then to be able to go and find the really good interesting stories that can actually break through and have a larger audience care about them, that's tough. To have a newsroom that's going to give you the time to do that, also difficult. Everybody in newsrooms are overloaded. Journalists, about 2000 media jobs have disappeared in the last couple of weeks in North America. So I think there's that tricky element of the technical stuff and the time to get up to speed, and then also making people care about it. How do you talk about this in a way where people care the fact that hundreds of millions of dollars might have been stolen from brands? Brands aren't the most sympathetic thing in the world. So those are all reasons why for coverage choices people don't focus on it.

Then I think the other pieces that of course we have an industry press, a trade press that covers marketing and advertising and you would think they would be the ones to do this, but they don't tend to do investigative work. They don't tend to really dig in and reveal the CD stuff going on in an industry, and I think frankly of course they have to run events as part of their business models. They need to have relationships with vendors. So if they're going out there and calling people out and making things uncomfortable for vendors and other people in the industry, well that's a bad thing for their business.

So unfortunately I think there're a few factors that have made it so that journalists aren't really looking into it, and so that level of scrutiny isn't getting applied on top of some of the industry things that happened.

[00:08:52] JM: What are some of the ways that advertising fraud manifests?

[00:08:58] CS: There are a lot of different ways, and one of the things that I've come to understand is that people – Whatever new technology is rolled out, whatever new ways people can bid on ads, whatever new advertising formats, fraud follows that, and if the money flows into it, fraud follows that.

So on a basic level, you can have scenarios where somebody sets up a website where they copy content or buy some cheap content, put that on it, and then they either create, say, bots or

pay for bots to go and visit the pages where they have ads and they earn money. So that's a basic kind of ad fraud where you've created kind of fake or low-quality media and then you've created or purchased a simulated audience to make sure those ads gets feud. So that's on a really simple level of it.

Stuff that's a little more complex is in these real-time bidding exchanges where a website or an app is making inventory available for brands to bid on and place their ads at, sometimes people are going in and they're pretending to be, say, buzzfeed.com or they're pretending to be the Wall Street Journal and somebody thinks they've bought an ad that is running on one of those known properties, but instead it's actually just running on a completely fabricated website. So that's another way where it's called spoofing, where people are pretending to be real media and pocketing the dollars.

So from fake media and fake audience, you get so many different kinds of permutations. Another one that I wrote about recently, just to pick another type out, is called attribution fraud. So a lot of mobile app makers, they want to drive installations of their app. So they go out and they'll pay what's called a bounty. So if you through your advertisements get somebody to install, say, Uber, then Uber is going to pay you maybe $2, $3, something like that for that. So you have a bunch of apps that once they're on somebody's phone, they're looking to see if people are downloading other apps and they're injecting their own kind of attribution for it saying, "Hey, I caused this person to download Uber on to their phone. You need to pay me out," and this is a rampant practice.

The story I did recently was about a huge Chinese developer of mobile apps called Cheetah Mobile, which several of their apps, according to the data that we were shared, were engaged in that kind of attribution fraud and Google actually ended up removing one of their apps completely from the store and that investigation continues.

[00:11:27] JM: When ad fraud occurs, where's the money actually being stolen from? Who is losing in this equation?

[00:11:35] CS: This is one of the things that is kind of wild about it. I mean, one the ad fraud is occurring because there are so many different middleman in the placing of an ad. There is an

example done recently where people try to kind of figure out the way an ad got placed on a well-known website, the Business Insider, and how many people touched it on its way there. It's incredible how many people are trying to take their cut and their piece, and that complexity is what creates these opportunities for the fraudsters.

But at the end of the day, even if there are 5, 6, 7 different hands, 7 different middleman from somebody trying to place an ad till it actually ending up on the site, at the end of the day, the people who are losing the money are the advertisers, are the brands. So if you're a company like P&G who's spending one of the biggest advertisers in the world, you have to think if you're going to spend a billion on advertising, you're probably going to lose somewhere between 15% and 30% of that to fraud, and you're not necessarily going to know that unless you go and you are rigorously looking through all of the places that your ad showed up.

So you can imagine, that's a Herculean task for an organization to do. But at the end of the day, it's the brands. The brands are losing their money. It's getting stolen because they're the ones spending to try and reach people.

[00:12:49] JM: Right. So I think one analogy we could draw is how in the 2008 housing crisis there was a huge network of different securities that were being sold and repackaged, and resold and repackaged, but ultimately the person who got hung out to dry was the owner of the real estate who had taken on some sort of debt and they really got damaged. The same thing is happening to the brands, where you have all these middlemen of ad agencies and ad networks and we would need an audiobook length podcast to actually dissect the spaghetti of different players in the advertising world, but the end result is that the brand is the one who is suffering. So why aren't the brands rebelling at this?

[00:13:40] CS: Yeah. So to a certain extent they are, and I think we've seen really over the last 2-1/2 years brands demanding more accountability for where their money goes. But it is wild to think of how many years they just sort of accepted what was being told by them by their agency or the other partners they were working with, that, "Oh, yes, you met your campaign goals. Congratulations. Everything went well." So they are agitating for more.

There’s an industry initiative called ads.txt, which is requiring websites and soon, hopefully, mobile apps to kind of disclose which networks they work with so that people can compare and not have spoofing happen. There are other industry initiatives that are happening, and then you have people like for example the chief marketing officer of Procter & Gamble did give a speech about a year and a half ago where he talked about the murkiness of the industry, where he talked about the theft that was going on. So they’re agitating.

But here's the thing about this, this industry, the digital marketing, digital advertising industry, is set up in a way where there's so many incentives for people not to report fraud for people to continue to allow it to happen. It is so completely messed up, and you would wonder, “Okay. No brand wants to lose their money. Obviously, they want to reach the right people.”

But let me give an example here. So I did a story at the end of 2017 and I looked at a bunch of seemingly reputable digital publishers who had all decided to go out and buy audience. So if you're a reputable digital publisher, you're not supposed to be doing that. You are not supposed be paying somebody to funnel audience to you. You're supposed to have built a brand of people come to maybe from your Facebook page or other things.

So they had all decided to buy this cheap traffic to sort of see if it worked on their site and if these people they were buying came back, and the traffic turned out to be fraudulent. So we exposed that and we exposed the publishers that had been buying it, and in one case, one of those publishers had bought this traffic and directed it to the sponsor content from a major bank. So you would have to think the bank would not be happy about this. The bank would go to that publisher and say, “Wait a second, we paid for real people to view these sponsored stories, and instead you go out and you buy cheap traffic and you throw it at it to meet our metrics, to meet the campaign goals. So I informed the bank of this. They were very grateful. They looked into it, but at the end of the day they didn't actually want to publicly say, “Yes, our partner had ripped us off here.” They didn't want to do that. They gave me a statement where they actually claimed that the sponsor content that I identified having received this traffic wasn't actually part of their agreement.

So this is a case where clearly something completely improper had happened, but they publicly do not want to call it their partner. I think the reason is that if you’re that CMO, then you’ve got to

explain to your CEO why you entered into a partnership that ended up like this, and then they're naturally going to question, "Well, how much of your other spent is being wasted? How much of your other spent is going to fraud or misleading things?"

So everybody wants the budgets to continue. Everybody wants to get their big-budget from the board and from the CEO and they want to be able to spend it however they can. If suddenly everybody loses confidence in the ecosystem, whether it's boards and CEOs to other people, then the money is going to dry out. So even the brands themselves, I have seen cases where they don't really want to rock the boat on this. I mean, I've heard that in that case, there was obviously some negotiation and talk behind the scenes, but in terms of a public calling out of this, they didn't want to do it. I think that's a really big problem that exists in the industry right now.

[00:17:23] JM: To present a another tortured analogy, sometimes think of this as kind of like climate change where you see, and not to get into like a debate about climate change, but you see these isolated incidents where you have gigantic fires in California or you have these tremendous storms that are occurring and you look at these things, you're like, "This seems abnormal and this seems somewhat disconcerting," but it's really hard to quantify what is going on in the macro picture with regards to climate change, or forget if it's humans are not or whatever, just like what's going on?

Advertising fraud kind of feels like that, where you see these isolated incidences where like the reporting that you do, which is why it's so important, you uncover these gigantic fraud schemes that are conducted by like four people and you're like, "Oh my God! This is concerning," but you have no way of knowing the scope of it, because how would you even crawl the internet for the scope of it? That's why you see these estimations like, "Oh, $19 billion is going to be lost to advertising fraud," and these are such unscientific quantifications. Do we have any way of knowing how much money is being lost to advertising fraud?

[00:18:40] CS: I think that nobody knows the real number. It's wild, isn't it? It's such a crazy thing, because the promise of digital marketing was supposed to be the ability to measure, the ability to reach the right person at the right time with the right message and know whether you reached them. All of the tracking, all of the analytics, this was supposed to usher in a new age of

marketing, and instead we have just oceans of garbage data and an insane amount of fraud. I think the promise of digital media has been squandered.

So do we know how much is being stolen? No. I talk to people all the time and what's the percentage of digital ad spend that's being stolen by fraud? Well, if you talk to people in the industry who have an interesting kind of making it seem like it's not such a big deal, they talk about numbers where it's like, "Oh, it's maybe 5% or less than 5%, or 10% at the most," and they put out their numbers to reflect that.

Then you have other people, you have vendors who sell technology that they say, "We'll protect you," the brand, "from fraud," and of course they cite extremely big numbers.

So everybody has their own self-interest. Everybody's got their own statistics. Everybody's got their own methodology, and at the end of the day I do have some sympathy for the brands and the agencies and others because you're just getting whiplash from all the different takes and all the different people with different interests, self-interest, who are just messing with you. So I think that's a huge problem.

The climate change analogy is probably a good one. I do like your financial analogy. I read Flash Boys recently, which is of course the story about high-frequency trading, and a lot of that really resonated with me when I looked at what was being laid out there, where what was going on in high-frequency trading was you had people who really understood how the technical elements of stock exchanges worked.

To give it a really quick summary, I mean, they realized that if they positioned their servers for making orders and for watching the orders happening on the exchange, if they position their servers close to the ones of the actual stock exchanges, well they can get information before other people. They could act on it and then they could basically take money from people, because they saw stuff happening before other people saw it happen.

I think that is a lot of what's going on in ad fraud. You have people who understand the technical elements of ad exchanges, of programmatic ad buys, of all the different things going on, and instead of using that knowledge to be a good player, they use it to exploit and they add no

value. They're just taking money out of the system. They're exploiting the knowledge and the understanding they have and they're exploiting opaque and how many middlemen and how many technical layers there are in that infrastructure. I think that to me, reading that book really resonated. I think the cancer that high-frequency trading was for the stock market, we're seeing that with ad fraud, where insiders are really exploiting their knowledge. People who think they're really savvy, people who think they're really savvy marketers are the ones who are getting exploited, just like traders at well-known banks were getting exploited because they didn't realize that their trades were being seen by these system, these algorithms before it was getting to the rest of the market.

[00:21:50] JM: This is of course why many of the biggest ad exchanges are based in New York.

[00:21:57] CS: Yeah. Yeah, the infrastructure, that's there. It's not a coincidence I think that you see that happening as well, and there's other pockets of kind of ad tech as well. The big investigation that you mentioned at the top, it turned out that at least a couple of the people who were at the top of that fraud scheme were actually based in Israel, and there's a huge amount of ad tech companies in Israel. So wherever you have a cluster of ad tech companies, you have a cluster of fraud happening as well, because that's where the expertise is and that's where all the buys and the sells are happening.

[SPONSOR MESSAGE]

[00:22:38] JM: DigitalOcean is a reliable, easy to use cloud provider. I've used DigitalOcean for years whenever I want to get an application off the ground quickly, and I've always loved the focus on user experience, the great documentation and the simple user interface. More and more people are finding out about DigitalOcean and realizing that DigitalOcean is perfect for their application workloads.

This year, DigitalOcean is making that even easier with new node types. A $15 flexible droplet that can mix and match different configurations of CPU and RAM to get the perfect amount of resources for your application. There are also CPU optimized droplets, perfect for highly active frontend servers or CICD workloads, and running on the cloud can get expensive, which is why DigitalOcean makes it easy to choose the right size instance. The prices on standard instances

have gone down too. You can check out all their new deals by going to do.co/sedaily, and as a bonus to our listeners, you will get $100 in credit to use over 60 days. That's a lot of money to experiment with. You can make a hundred dollars go pretty far on DigitalOcean. You can use the credit for hosting, or infrastructure, and that includes load balancers, object storage. DigitalOcean Spaces is a great new product that provides object storage, of course, computation.

Get your free $100 credit at do.co/sedaily, and thanks to DigitalOcean for being a sponsor. The cofounder of DigitalOcean, Moisey Uretsky, was one of the first people I interviewed, and his interview was really inspirational for me. So I've always thought of DigitalOcean as a pretty inspirational company. So thank you, DigitalOcean.

[INTERVIEW CONTINED]

[00:24:45] JM: Now, because you are one of the – Actually, I think I'll be completely honest. I think you are the best journalist at covering this topic just because of the depth that you have crawled down into the muck of this stuff, and I've tried to cover it in the podcast and just like people don't really seem to care or they don't really seem to understand just the ramifications of what this suggests about the realities of our internet. But because you have crawled down those steps, I want to break at the tinfoil hats a little bit earlier in this episode than I would normally. What role do you think Facebook and Google have in this ad tech world?

[00:25:27] CS: Yeah. Well, I mean, these are the 800 pound gorillas or however you want to talk about it. Now they Slightly different roles. I mean, let's start with Google here. So Google is the most dominant player in all of digital advertising. In every single part of you think about the tech stack of digital advertising, Google is a player in basically all of them. If you want to make inventory available for other people to buy if, you can do that through Google. If you want to go out and buy inventory, you can do that through Google. If you want a place ads, you can do – I mean, it's just nonstop, and it even goes into the mobile world where of course Google created the Android operating system and Google has the dominant Android App Store, the play store.

Google also has a dominant mobile ads, SDK, where if you want to get ads in your mobile app, well, you can work with Google to do that. So they are everywhere. They're up and down the

chain, and of course if you accept that fraud is a reality, which everyone does, they disagree on the depths that it goes to. But if you accept that fraud is a reality and you understand that everybody along the process from a brand deciding to spend his money to that ad getting placed in an app or on a website, everybody through that process makes money.

If you realize that Google can potentially be at every step of that process, well, then you realize that Google does make money from fraud 100%, and Google has teams dedicated to ferreting it out. They've done some good work exposing some things, but there is a tension at the core of Google's operation, whereby stamping out fraud would actually take billions and billions out of the ecosystem and probably would affect their bottom line.

So one of the discussions in the industry that happens a lot is Google tried to walk this line of caring about fraud and trying to stamp it out, but also at the same time making money from it. There's where Google sits, and I mean it is a huge, huge impact that they have on digital advertising, which is why senators like Mark Warner are really looking at that and saying like, "Do they have too dominant in position?"

Facebook, a little bit different. Facebook, obviously, people buy ads on Facebook, and those ads I think for the most part, Facebook is a pretty effective advertising platform. Facebook also has an ad network with people off of Facebook. So similar to Google and AdSense, with Facebook, you can sign up to Facebook audience network and you can monetize your app and you can monetize your website with them.

So there we have Facebook with dominant platforms where the people are where they can serve ads to them. Facebook also has different ad products where even if you're not on Facebook, you can help earn money through it. So, again, Facebook earns revenue from fraud. Anybody who is earning money from digital ads being placed is earning money from fraud. So they have a huge role and they both talk a very good game publicly about caring about it, but they are in a conflicted position, because if fraud was completely wiped out, that would affect both of their bottom lines and it's a really tricky position and it reinforces how the incentives in this industry are all screwed up.

[00:28:39] JM: And one flavor in which this is not like the 2008 crisis, is that there is so much legitimate value being created by these advertising systems. I mean, ads really work for some businesses in some contexts, but one thing that you cover in your article about We Purchase Apps and fly apps and this network of four people that made between probably $75 and $750 million, something like that, is that there is this – I call it traffic laundering, where basically you have like some legitimate traffic and then you also have some bot traffic and you train the bot traffic to look like the legitimate traffic such that it becomes almost indiscernible who is real and who is fake and you funnel so much real and fake traffic through the systems that it becomes impossible to discern who is real and who is fake.

[00:29:35] CS: I mean, it's a wild thing. This specific attack, people often refer to it as a recorded attack, where they're recording the behavior of actual real human users on a website, in a mobile app, and then once they've gathered that and they understand where the user is coming from. What time of day are they in the app? How long do they spend? Where do they tend to click? All these things.

One, realize that when you're using an app, you've probably given permission for them to completely record what you're doing. So people should be aware of that. Then second, because of that, because of the permissions that we handover, or on the case of a website, because of the good open protocols on the web, people are able to really get down to the nitty-gritty and unsavory folks who've been doing these recorded attacks. They're not entirely new, but the sophistication of them continues to advance, and it is such a freaky thing to think that a real human audience was tracked and clone and then that exact behavior was programmed into bots to then go and load those apps.

If you think about the ramifications of that, the goal here is, one, to of course increase their audience, which increases the amount of ad revenue they can get. But two, it's to bypass the fraud technology, the fraud section technology, that is built-in. This exposes one of the flaws and a lot of the fraud vendors, antifraud vendors, that are out there where they're taking kind of a data science approach where they're sampling the traffic and saying, "Okay, here's what her baseline looks like, and now, wait a second, we've just got new traffic that is way off of the baseline. This looks like fraud to me."

So if on the fundamental basis, the fraud, the bots are based on real human behavior, you're going to be able to bypass a lot of the fraud detection that is out there and it shows how clever and smart a lot of the fraudsters are. It shows how they understand exactly how fraud detection is done and they build their systems to get around that and to evade that.

I think the larger thing here taking out of the realm of ad fraud for a second is that people need to realize the amount of real traffic versus pay traffic on the internet as a whole. I mean, there are some days where there's a lot more fake traffic than real traffic. This is the world that we've built and it's really a profound and mind-boggling thing to realize that on a lot of days of the year there's more bot traffic than actual human traffic on the, and the implications for that in advertising are massive, but also for other parts of the things we do every day.

[00:32:06] JM: Yes, I am so glad to have found somebody with a tinfoil hat size that is the same as mine. This is like why – Like I've tried to do interviews with some of these bot detection companies. If you look into the episode history, I have done one or two shows with bot detection companies, and you get to a point in the conversation where you say, "Okay. So how do you detect who is real and who is fake?" and they're like, "Oh, well, we have this combination of signals and blah-blah-blah-blah-blah," and I'm like, "Okay. So if somebody makes it through your system, how do you know that's not just like a false negative or a false positive or however you would want to put it." They're like, "Well, we really can't tell." I'm like, "So how do you make these guarantees? Do you catch 80% of bots?" and they're like, "That's what we think we catch," and you hit this point of circular arguments, and it's crazy. It like a technological dogma where you hit these circular argument and it's just like – It's like are we living in the same universe? Have you countered these people with this cognitive dissonance?

[00:33:12] CS: Yeah, I mean if you, like me, and your inbox is filled with people putting out press releases about fraud detection, technology and stuff like that, I mean, you see so much garbage, you see so many ridiculous claims that are coming through. So you have a real snake oil problem. That's not to say every fraud detection vendor is snake oil. I mean, on a lot of the stories I've done, I've gone to these companies to see what data they have to compare it against what I'm finding, and this is actually a really telling thing for me, is that on any given story there may be a couple of fraud detection vendors who are really helpful who say, "Yes, a

year ago we saw that this traffic was coming. We blocked it. Here's what we saw from them," and they can often share stuff that you can sort of validate and put up against what you have.

But there were also cases where, and this is another huge problem in the industry, there are also cases where I might be investigating, say, a particular publisher, and that publisher has paid one of these fraud detection vendors to kind of verify and validate its traffic to say, "Yeah, yeah. This site has 95% real human traffic," and then advertisers feel safe buying it.

Well, if I find something that shows there's a problem with that publisher and one of these fraud detection vendors is basically verifying them and saying they're clean, that vendor wants nothing to do with me. So on any given story, a fraud detection vendor can be extremely helpful to me, or because they're being paid by a particular party, they clam up and they won't talk to me. This is a huge problem where these vendors will take money from everybody in the ecosystem. You can be a brand and pay to validate site or the ad slot you want to buy. You can be a publisher and pay different people to say that your stuff looks really good. As long as everybody will take money from everybody, there's no one who's sitting there and willing to actually call it out.

So there's some problems with the methodology and some companies that are absolutely snake oil, but even a lot of the good ones, they're taking money from everyone. So on any given moment, they may actually be saying a complete garbage property is totally fine.

[00:35:19] JM: So one thing that I have seen is, is I see incentives in place that make it such that Twitter or Facebook would actually want bots on their platform, because bots consume ads. Is that too paranoid to think?

[00:35:40] CS: Well, I mean, I think of course they would prefer to have real human users, right? That's what they would love to have, but if you're in a case where – And I think it's less of a problem for Facebook and more of a problem for Twitter, but if you're in a case where growth is tough and you're up against huge behemoths with far more people, then the tendency to want to grade something as an acceptable account that might be unacceptable somewhere else, I think that's there, and I think Twitter as we all know has a massive bot problem. It's only recently

been starting to address it in any kind of a meaningful way, but at the end of the day, Twitter wants to have as big an audience as possible.

So what they consider a bot or an automated account, a policy violation, and what other people might, there's probably a big gulf between that. The reality, of course, is that bots do view ads, and if they want people viewing ads, they're going to earn more money if there are more bots viewing it. Again, yeah, we have this incentive problem where I know there are people at these platforms who every single day wake up and try to rid this stuff from there. There's no question about that, but then there's the business case.

The business case is get away with what you can, because your revenue is going to grow, and there is a big debate happening right now about the numbers that Facebook releases, the numbers that Twitter releases where they estimate how many fake accounts or bots or what have you are on their platforms, and the numbers, especially with Facebook, the numbers have been growing a huge amount and yet Facebook isn't really acknowledging that the problem is getting worse.

With Twitter, people have always suspected that their monthly active user numbers, the stats that they put out, have included a lot of really garbage accounts or automated accounts. So there's a lot of skepticism about them. Again, yeah, we've got the scenario where I think they would all love to have real human users completely, but also there is an incentive to allow a certain amount of bots to exist. There is a strong financial incentive.

[00:37:40] JM: Well, what you say about Twitter versus Facebook, I kind of think that the only reason we think there are more bots on twitter than Facebook is that Twitter is this open world where you don't have as much filter bubbling, but like in Facebook, you have these isolated communities, like people don't actually see everyone else on the platform. That's why Twitter is more fun in many cases, but Facebook has these isolated communities of like people where if you don't have a close connection to them, like if there's 18° of separation between you and a bot, you're never going to see that bot. So we don't really see the full picture of Facebook as often.

[00:38:18] CS: I think that's true, yeah, and this is one of the biases that journalists and a lot of researchers have, which is that Twitter is a more open platform. You can get more information from its API. It's better for researching. It's easier to research. So, yeah, we have this kind of level of scrutiny on it that is probably unwarranted given the amount of people who are on it compared to Facebook.

Facebook is much more of a black box and they're actually becoming more and more opaque. Because of the Cambridge analytics scandal, they're turning off a lot of the APIs and other things that were previously available and open to you. So it is hard, and it's also I think a lot harder to identify a bot on Facebook than a bot on Twitter, and I've worked at doing both.

I went down a crazy rabbit hole last summer where I thought I had come across a really massive botnet pushing out hyper-partisan political content. But when I got down to the user level and started to figure out if these belong to real people or not, the crazy thing was that I would figure out that there was a real person with this name in this location. In some cases, I had email exchanges with them or phone calls with them where it was usually older people who are like, "Yea, I just use Facebook a lot," and these were folks who would in the span of five minutes just hit the re-share button like 30 times. It looked like automated behavior, but they were actually real users.

There was something about behavior on Facebook I think, and I'm trying to really dig in to this more particularly around older users where they're behaving in ways that do not seem human. It's really weird.

[00:39:53] JM: I know. I've seen it – And I say it that way because I've –

[00:39:57] CS: I think a lot of people can relate, yeah.

[00:39:59] JM: Literally, like I've seen it in relatives.

[00:40:02] CS: Yeah, that's it.

[00:40:02] JM: You see your older relatives sharing this insane stuff and you're like, "Aren't you the elder? Don't you have like a sense of this being completely fabricated?" and they don't, because they didn't grow up on the internet.

[00:40:02] CS: Yeah, and I think a lot of us are struggling to adapt to this new media world. It really is different, and as much as I'm kind of interested in trying to figure out how much age is a factor or not, I think all of us are having cognitive difficulties at times to really deal with these streams of information and all the sources and information overload, but it does seem like particularly for older folks who are trying to grapple with this, there are certain behaviors and certain things that makes it seem like it's a particular difficulty or even more difficult for them.

I would also caution against assuming that younger people are always naturally great with this stuff. I think they too need some of the skills to navigate this world, and it's going to take us a long time to adapt to this. I think that's the bottom line for me is it's not just like you get used to using Facebook or you're getting used to using Twitter. It's a very different way of consuming information and the consequences of that I think are still yet to play out.

[SPONSOR MESSAGE]

[00:41:25] JM: This podcast is brought to you by wix.com. Build your website quickly with Wix. Wix code unites design features with advanced code capabilities, so you can build data-driven websites and professional web apps very quickly. You can store and manage unlimited data, you can create hundreds of dynamic pages, you can add repeating layouts, make custom forms, call external APIs and take full control of your sites functionality using Wix Code APIs and your own JavaScript. You don't need HTML or CSS.

With Wix codes, built-in database and IDE, you've got one click deployment that instantly updates all the content on your site and everything is SEO friendly. What about security and hosting and maintenance? Wix has you covered, so you can spend more time focusing on yourself and your clients.

If you're not a developer, it's not a problem. There's plenty that you can do without writing a lot of code, although of course if you are a developer, then you can do much more. You can explore

all the resources on the Wix Code's site to learn more about web development wherever you are in your developer career. You can discover video tutorials, articles, code snippets, API references and a lively forum where you can get advanced tips from Wix Code experts.

Check it out for yourself at wicks.com/sed. That's wix.com/sed. You can get 10% off your premium plan while developing a website quickly for the web. To get that 10% off the premium plan and support Software Engineering Daily, go to wix.com/sed and see what you can do with Wix Code today.

[INTERVIEW CONTINUED]

[00:43:24] JM: Now, we've been talking mostly about logged in experiences. So if you're on Facebook or you're on Twitter or even kind of, I guess, on Android in many cases, you're engaging in a logged in experience where there is at least some identity information associated with you, but there are these other cases where it is more of a logged out experience. So like when ad fraud takes place on the open web, it's even more of a Wild West, because there's kind of less signal for these companies to potentially filter out who is a bot and who is a human.

In this fraud investigation that you did, one part of it was you – These set of fraudsters, and I know we're not really delving down into the story as much as I would've liked. Maybe we'll talk about that a little bit later, but people can certainly read the story that went viral that you wrote, and I'll include it in the show notes, but the fraudsters, they operated these several web properties where they had fake video advertising. Can you describe how the fake video advertising on the websites? Like just to give people another perspective into how one of the schemes works. Describe the fake video advertising websites.

[00:44:41] CS: There's kind of two flavors actually that have come out recently. So one, in this story I wrote where the guys, they had like over 125 android apps and they also had these websites, as you mentioned, where basically these were websites where there was like nothing there. They had just licensed the content from some content providers, just put them and they copied and pasted the same about text on every website or they plagiarized it from elsewhere. There wasn't really a real audience on them. I think what happened is they bought the audience and then cloned it and were showing them video ads.

The reason for video ads being such a big thing is, of course, the CPM's are higher. You earn more money for showing a video to someone than you do showing a banner ad. In the case of the video stuff, I mean, these were just video plays that were being logged as having happens. The users looked real enough that it was being allowed by the fraud detection companies, and these guys were just racking up insane amounts of video views.

Some of these video sites where if looked at them you would realize no one would go to this site. No human would go to this site, because there is really nothing there except this really silly generic celeb content. Some of them had insane amounts of available ad inventory, meaning that they were claiming to be serving millions of impressions a day to users. This sort of exposes one of the fundamental problems where you have all of these brands deciding to pay for data science-driven detection, but literally if they sent a human to visit these sites, they would realize they would never want to advertise on them. But because so much is done automatically and programmatically and because the audience looks great because it's engineered to look great, this stuff goes through.

The second example here aside from what those fraudsters were doing, there was a huge fraud scheme called Methbot, which is actually resulted in multiple indictments and two guys being extradited to the United States, and they're trying to extradite some others, but one of them is in Crimea. He is not getting extradited.

So in this case, these guys actually built kind of a fake web browser. So a headless web browser, which is often used for testing and other things. They built a web browser that would just make itself appear in the ad systems as if it was a real website with a real audience serving up tons of video, and they basically just created this fake web browser that they loaded on tons of different data centers and ran like crazy and it just ran tons and tons of video ads and they earned millions and millions of dollars. So they didn't even actually need to setup much of a website, even less of a project than the other fraudsters, they really just built this clever web browser that fabricated everything about an audience and everything about a website so that people would actually think they were getting ads in front of real people. It's wild. If you have the technical expertise, the things that you can get away with are really remarkable, and I say that is a cautionary note, not as a how-to to anybody listening to this.

[00:47:42] JM: Right. Just to give a little bit more context on this business that you covered, this four person, at least four – Do you think it was more people or do you think it was just basically four guys that were running this business?

[00:47:55] CS: So there were four guys at the kind of the top of the pyramid. Two of whom had a background in digital advertising and two of whom had actually run like an ISP in Germany. So you can see how the technical expertise came together, but in order to actually execute the fraud, they did at times hired different people to do things like go out and license the content for those video websites. They worked with people to go out and acquire real apps that they can then move into the fraud scheme.

So the four people were the core of it, but then they would you have contractors and other people working with them at different times, but the design of the system and maintenance of the business really came down to four people to steal potentially hundreds of millions of dollars. I mean, if you think about the amount of people and money they needed to invest in this to get up and running in order to earn that much, I mean, that's a pretty good business. That's the problem. Fraud is a really good business.

[00:48:48] JM: De minimis. It's easier to start and ad fraud company than to start a startup, basically, to my mind.

[00:48:55] CS: Absolutely.

[00:48:56] JM: So not to give any of the entrepreneurial listeners any ideas. One think I liked about the story is how much of a story it was. It mean, it opens with this person who gets approached by this company. I think it's We Buy Apps or We Purchase Apps. Imagine you're a developer and you make some random app in the app store and it catches some virality. It's like you create the next flappy bird and somebody comes to you – Or you create even something that's much worse than flappy bird, like a workout app, like this is your ab crunch app. Then somebody comes to you and says, "Hey, we see you. You're kind of popular. You've got like a thousand daily active users. We want to buy your app," and you're like, "All right, sure. How

much are you going to pay me?" Then they offer you much more than you anticipated and you're like, "Well, yeah. Sure. I'll sell you my app."

So We Purchase Apps buys your app and then they take your thousand users and they record the behavior of those 1,000 users and then they spin up your 50,000 bots that replicate the thousand real user behavior and they sell a bunch of ad inventory on the ad. Is that how it works?

[00:50:05] CS: I mean, that was a good – It in a nut shell, and the interesting thing is each these apps that they were buying, none of them were, in most cases, with one exception, none of them were particularly like huge big hits. In fact, that was what they targeted. They wanted apps that managed to build like a small, fairly loyal following that look like real people that had positive reviews that had been around for long enough that it had a bit of a reputation in the play store and another places. Then they would approach these people.

So these were apps that they could buy for 10 grand, 15 grand, whatever, and the person who created it would be really happy, because for them it was a small little business and then they have to maintain the thing, and now they get a big cash payout and they walk away. Whereas for these guys, once they got the apps, they had an operation in Serbia where they had Android developers, tons of people, who could then maintain the apps, who could update them, put in the right ads, networks and things that they work with. So they started to create this kind of scale where they had dozens and dozens of apps.

In a lot of cases, I had an analyst from Malwarebytes analyze a bunch of the apps and he found that that they started the kind of create templates within the apps themselves. There would be the core functionality of what was often a game, but the rest of the stuff for ad serving and all of that, they pretty much clone them. So you can see how they created efficiencies, but it's wild for me to think you had the four people at the top, but they did actually have this business that had a website and looked like a legitimate business that did employ a bunch of developers and other people to maintain these apps, and in some cases they also develop new ones on their own.

So it was a wild operation to think that they came up with this solution of like, "Well, hey, rather than trying to build our own apps all the time and grow an audience, let's just pay money to get

apps with real audiences. Let's clone them," and then it's just like you turn up the dial. You just turn up the dial, you double the audience, you triple the audience, quadruple the audience, and this was one of the things that some people got on to them for.

So I actually did a follow-up story early this year about a company in Israel called Woobi that these fraudsters had approached and wanted to help place ads and their aps with. So Woobi did an initial due diligence. Everything looked okay. So they had these apps in there, and Woobi only works with games. But after a little while, I don't know, six months or so, Woobi started to be a little bit suspicious, because these apps that weren't very high-quality, that weren't really good games, suddenly their audience and the amount of ads that they were showing to their audience was really out of whack for their experience for other games and it got to a point where some of the fraud vendors were starting to flag some of the traffic in these apps. There was one fraud vendor in particular called Protected Media who sort of saw what was going on early on.

Woobi eventually told three different – I think it was three or four different companies that had these apps that it can't work with them anymore, because there's fraud in them. Woobi didn't actually know that all of these companies were connected. They were different shell companies that these guys had set up to put different apps in to go out and do ad deals, and the only way Woobi actually got on to things in the end, aside from the fraud, was they sent out a message saying, "Hey, we want to send you a gift for Christmas. Give us your address," and these four seemingly separate companies all sent the same address, which was basically – I think it was in Malta and it was just like a shopping center or a business center. It wasn't a real address, and they realize these four companies that they'd have some fraud problems with were all the same, and that's when they cut them off.

So they were smart in a lot of ways, but then they messed up and I was able to really expose them, because they had made mistakes like reusing the same about text on their websites, like reusing the same address for different shell companies and other things that we were able to connect all the stuff together. The end of that story with Woobi is that when Woobi kicked them out and said, "We're not working with you anymore," these guys actually sued them. So you have fraudsters who feel so emboldened that they will sue a company even though they've been caught with their fraudulent inventory, and that case was only resolved at the very end of

last year in part because of me having exposed what was going on. They didn't pursue the case anymore.

So, I mean, that to me as an example of the industry there where the lengths they go to to set all these up, but the fact that they would even file a lawsuit after somebody caught them and try to get money from them, like there's no fear. The fraudsters have no fear.

[00:54:37] JM: Just give people some context on these apps. I think when I scroll through the spreadsheet that you link to in your article of the apps, I mean, there is an element of humor to this whole story. That's actually one reason I really like covering this area, is because it's kind of ridiculous, but you look at these apps and it's like selfie expert pro plumber mania minesweeper, like somebody is playing minesweeper on their phone.

I mean, there are – Guess, the restaurant quiz, the logo trivia game, and it's like I know that you and I probably would never download these things. Maybe I would want the grumpy weather widget or the gluten-free food finder, but occasionally maybe you're at Thanksgiving and somebody like hands you their phone. Like on the rare occasion where you actually see somebody else's phone. I don't know about you, but occasionally I'm just like, "Why are you downloading this stuff? You don't want to download this. This is a bad idea." People just don't know. They don't get it, because it's like we remember back in the day when you download the toolbar and it starts sending you pop-ups and stuff, and those people are still around. The people who are making the pop-up toolbars are still around. They're just like a little bit more savvy, and that's what this stuff is.

[00:55:50] CS: That's literally the case with this scheme. Two of the guys ran toolbar companies before they moved into mobile apps. I mean, that's exactly what they did. They moved from that world of ad fraud to this one. I mean, the app scenario is really – I mean, I probably use the word insane a lot so far out, but what's going on particularly in the Android ecosystem is absolutely insane. The amount of malicious, invasive ad fraud committing apps that are in the Google Play Store with tons of downloads and out there and available to people is completely out of control. I mean, we see examples of this all the time. I mean, some of the worst offenders are flashlight apps. Every phone comes with a flashlight now, but somehow there are flashlight apps in the play store that claim to have like hundreds of millions in some cases of ad

impressions available every day for people to buy. As if people are using flashlights that much, and it's just such a clear example of fraud, but they're there and they're available in the store. They're available in places like even Google's own ad exchanges, and they're not being cleared out, and a lot of these apps are also taking insane levels of permission to the point where they can see everything you are doing on your phone and people are not paying attention to what they're downloading. They're not paying attention to permissions they're giving.

In a lot of cases, not to sound too much like I've put on an even bigger tinfoil hat, a lot of the biggest app companies are Chinese companies, and this is not an anti-Chinese people thing. The issue here is that the Chinese government requires companies to make their data available for security and intelligent reasons whenever the government asks for it.

So, you have literally hundreds of millions of people all around the world using Chinese made apps that take insane levels of permission that have huge amounts of data on you and that data is literally being shipped to China where the Chinese military and security and intelligence apparatus can have free access to it. I mean, this is what is going on. I'm not a conspiracy theorist. It's 100% happening.

[00:57:48] JM: How is it that there isn't anybody else reporting on this? It's so shocking.

[00:57:54] CS: The app stuff gets more attention, and China right now is a big topic, but I don't think the app element of China has really been appreciated. People are focused on Huawei and device and equipment makers. So I do think we're going to see more reporting on this this year, particularly on the China stuff.

[00:58:12] JM: So what about Mark Warner? So the senator who is been vocal about advertising fraud. This is one of the few senators who seem a little bit more technologically savvy. What's he said? What are his actions revealing?

[00:58:25] CS: So Mark Warner – I got to give him – And it was actually a letter that he and Sen. Schumer sent back in 2016, before everybody became obsessed with Russian disinformation tactics as a result of the 2016 election in the U.S. They sent a letter to the FTC saying, "We are concerned about the prevalence of digital advertising fraud. We are also

concerned about the dominant position of Google in it. You need to look into this," and then everybody kind of got distracted by other things.

But when I did my investigation of that big app fraud scheme in last fall, Mark Warner did send another letter to the FTC saying, "You need to look into this," and because Google was referenced very much in that story, he said, "and you really need to look at Google's dominant position." Then when I did the story about Cheetah Mobile, the Chinese app company that was committing ad fraud and also taking huge amounts of permissions and data, he again said that we need to consider Chinese mobile app companies a potential national security threat.

So he is really the only one in the U.S. government who has raised any kind of alarms about this, and unfortunately his efforts to get the FTC to take this on have completely gone nowhere. He's actually been very clear and vocal to me and saying that they've basically ignored his letters, they've dismissed him and he's unhappy about that. So we're going to see what happens this year whether there's going to be more scrutiny of Google, whether the FTC might look into that.

I mean, I think right now there's probably more of a chance that the FTC looks into Facebook than Google, but Mark Warner and his office seemed to be very focused on ad fraud and also on potential data and security risks from Chinese equipment and app makers. So I mean, I'm hopeful some movement might happen this year.

[01:00:05] JM: What are you working on now? What kind of stories?

[01:00:08] CS: So I do have some more ad fraud ones in the hopper. I mean, I could literally write about ad fraud every single day.

[01:00:14] JM: Join the club.

[01:00:15] CS: The challenge which I – Yeah, this is it. I mean, it's that big of a problem and it's not getting enough scrutiny that part of me wants to do that every day, but it's really hard to find stories, like we're a general site. We have an audience that is not a technical audience and that is not an ad industry audience. So I have to find examples where the average person hopefully

understands the stakes and feels like there's something affecting them. I'm continuing to look at Cheetah Mobile. I'm continuing to look a lot at the Chinese app companies and not just because of the security stuff that's going on, but also they've actually been quite successful. They have a lot of big apps in these stores and it's interesting to look at the permissions they take and that kind of thing.

I'm also very interested in just some of the scams that are out there, like – So a couple of weeks ago I wrote about how people are renting out their Facebook accounts. So why would somebody rent out a Facebook account? Well, people rent out their Facebook accounts to organizations who want to then use their ads account that you have attached to your Facebook account to run ads for often very shady products, like skincare scams, and penis pills and that kind of thing.

So I'm looking more into that of how average people are earning 20 bucks a week, hundred bucks a month, to give some random person total access to their Facebook account, and in some cases total root access to their laptop and their router for like 100 bucks a month in order to let these people run shady ads, and this is something that people are doing out there and it's insane. So looking at that and trying to figure out who's behind this is a thing right now.

[01:01:57] JM: Well, Craig Silverman, thank you for coming on the show. It's been really fun talking to you, and I could not be a bigger fan of your reporting.

[01:02:05] CS: Look, I really appreciate it, and I appreciate that you've done several episodes now and that you're helping to put some awareness out there as well. So, thank you.

[END OF INTERVIEW]

[01:02:15] JM: GoCD is a continuous delivery tool created by ThoughtWorks. It's open source and free to use, and GoCD has all the features you need for continuous delivery. Model your deployment pipelines without installing any plug-ins. Use the value stream map to visualize your end-to-end workflow, and if you use Kubernetes, GoCD is a natural fit to add continuous delivery to your project.

With GoCD running on Kubernetes, you define your build workflow and let GoCD provision and scale your infrastructure on-the-fly. GoCD agents use Kubernetes to scale as needed. Check out gocd.org/sedaily and learn about how you can get started. GoCD was built with the learnings of the ThoughtWorks engineering team who have talked about building the product in previous episodes of Software Engineering Daily, and it's great to see the continued progress on GoCD with the new Kubernetes integrations. You can check it out for yourself at gocd.org/sedaily.

Thank you so much to ThoughtWorks for being a longtime sponsor of Software Engineering Daily. We are proud to have ThoughtWorks and GoCD as sponsors of the show.

[END]