EPISODE 596

[INTRODUCTION]

[0:00:00.3] JM: Ad-blockers in the browser protect us from the most annoying marketing messages that the internet tries to serve to us, but we still pay a price for these ads. We pay the bandwidth costs of requesting these pages. Our browsers are slowed down by these extra requests. Pi-hole is a hardware-based ad blocker. Pi-hole access a DNS server for all of the traffic that makes its way onto your network. Pi-hole has a blacklist of all the URLs to block including tracking systems and ad networks.

Pi-hole stops these URLs from communicating with all the devices on your network, including your cell phone. Jacob Salmela started the Pi-hole project and he describes it as a black hole for advertiser traffic. In this episode we explain how traditional ad blocking in the browser works and how things are improved with a piece of dedicated hardware doing the ad blocking. This was also a useful review of the relationship between URLs, IP addresses, your home network and the broader internet.

I also want to announce that we are looking for writers for Software Engineering Daily. We want to bring in some new voices. We're focused on high-quality content about software that will stand the test of time. You're listening to content about software engineering, you probably read content about software engineering. If you want to write, go to softwareengineeringdaily.com/write to find out more. We're looking for part-time and full-time software journalists.

We want to explain technical concepts and tell the untold stories of the software world. We're looking for writers who want to produce a single piece of content and also people that want to produce a series, or want to produce a long-form piece of investigative journalism. We just launched a new design for softwareengineeringdaily.com, so if you'd like to work with us, go to softwareengineeringdaily.com/write. You can also send me an e-mail, jeff@softwareengineeringdaily com. I'd love to hear from you.

[SPONSOR MESSAGE]

[0:02:09.0] JM: The octopus, a sea creature known for its intelligence and flexibility. Octopus Deploy, a friendly deployment automation tool for deploying applications like .NET apps, Java apps and more. Ask any developer and they'll tell you that it's never fun pushing code at 5 p.m. on a Friday and then crossing your fingers hoping for the best. We've all been there. We've all done that. That's where Octopus Deploy comes into the picture.

Octopus Deploy is a friendly deployment automation tool taking over where your build or CI server ends. Use Octopus to promote releases on prem or to the cloud. Octopus integrates with your existing build pipeline, TFS and VSTS, Bamboo, Team City and Jenkins. It integrates with AWS, Azure and on-prem environments. You can reliably and repeatedly deploy your .NET and Java apps and more. If you can package it, Octopus can deploy it.

It's quick and easy to install and you can just go to octopus.com to trial Octopus free for 45 days. That's octopus.com, O-C-T-O-P-U-S.com.

[INTERVIEW]

[0:03:40.4] JM: Jacob Salmela is the creator of Pi-hole; a piece of hardware that blocks ads. Jacob, welcome to Software Engineering Daily.

[0:03:48.4] JS: Yes. Thank you for having me.

[0:03:49.8] JM: You created a hardware ad blocker. There are plenty of software ad blocking solutions. Why isn't a software-based ad blocker sufficient?

[0:04:01.9] JS: Well, a hardware is nice because you can have a dedicated device, like a router would be, but it also just blocks ads for all your devices without having to install software on each individual computer, so you just install it once.

[0:04:15.6] JM: Are there any other problems with software ad blockers?

[0:04:18.5] JS: They're slow, because they try to use your computer's resources and load or block. They have to actually load the ad and then block it. Whereas Pi-hole will block it before it even reaches your network.

[0:04:29.5] JM: Why did you start working on an ad blocker?

[0:04:31.9] JS: I had a boring job, and I had a lot of free time on my hands, so I just started surfing the net and I got really sick of those advertisements. Then I backed this project on Kickstarter called the Ad Trap. It was basically the same thing as a Pi-hole is, but it was really not very well-made and it was slow and clunky.

[0:04:54.1] JM: What were the flaws with it?

[0:04:55.4] JS: Well, one of them was that it was just really slow. I mean, it blocked the ads, but then your whole network crawled through a slow, slow pace.

[0:05:04.1] JM: Was the software open source? Did you know what they were doing wrong on a networking level?

[0:05:08.1] JS: Yes. They were basically using a proxy, whereas the Pi-hole is a DNS server.

[0:05:13.1] JM: What's the difference between those two?

[0:05:14.6] JS: A proxy will route all your traffic through it, and a DNS server will just route only the DNS queries.

[0:05:21.3] JM: Okay. Explain what a DNS query is.

[0:05:24.1] JS: When your computer wants to look up where a website is, it's basically the phonebook of the internet. If you want to find out where google.com is, you have to send a DNS query to a DNS server and it will tell you where to find it.

[0:05:38.8] JM: Meaning it maps the URL to the IP address?

[0:05:43.7] JS: Yes.

[0:05:44.6] JM: Okay. Let's start from the basic premise of I go to a website with an ad, because I want to – explain this from a high-level view first. Let's say I go to beefrecipes.com. Actually beefrecipes.com is for sale, so if anybody wants to buy beef recipes, I checked before the show it's available, but I always use it as an example because it sounds like a site where you would go and you would be confronted with all these horrendous ads among beef recipes, like a recipe sites are some of the most atrocious ad-based websites that I've seen.

Let's say beefrecipes.com is this atrocious ad-based website. I go there, let's say I do not have an ad blocker and I load beefrecipes.com, there are banner ads all across it. What's happening there? When I enter beefrecipes.com into my browser and I don't have an ad blocker, I'm not sophisticated at all, all these banner ads are going to load. What are some of the different things that are going to happen as that page gets rendered and loaded?

[0:06:48.3] JS: Yeah, so you might actually only see that you're loading beefrecipes.com, but in the background there's actually a lot of other domains that are being queried, like Google ad domains, other analytics sites and stuff. Even though you're only visiting beefrecipes.com, there might be dozens of other domains being loaded in the background. That's usually how the ads are served, because it's easier to use a third party site than to dynamically host them on your own site, so you just use a service.

[0:07:17.4] JM: These ad networks that are getting contacted, is that problematic? It seems like they would be asynchronous with the page, and so it wouldn't actually hurt my pages' performance.

[0:07:28.0] JS: Well, a lot of times the ads actually load before the actual website content.

[0:07:32.3] JM: Interesting. The content just gets completely blocked until the ads get loaded.

[0:07:36.7] JS: Yeah.

[0:07:37.7] JM: What are some of the other types of companies that are getting contact aside from ad networks that tracking companies and other middleware companies? How much do you know about this soup of ad tech companies?

[0:07:50.9] JS: Just what I've seen when browsing the logs that Pi-hole generates, there's all sorts of different analytic companies, and sometimes even the company itself will be making calls home to different subdomains that they own.

[0:08:05.6] JM: Do you have any idea what these different companies do? I don't know if you've seen this – have you seen this loom escape thing?

[0:08:14.5] JS: No.

[0:08:15.1] JM: The loom escape, you should look this up some time, it is this graphic of all the different companies that are in the ad tech landscape and how they interact with each other and yeah, there's just thousands of companies in how they interact with each other is fascinating. The end result is that you have a lot of scripts on your page, of course, and they're doing all kinds of different things.

I guess, it doesn't really even need to be in your purview. You don't need to know what these different analytics companies are doing. You just know that they are all asymptoting towards doing the same work, which is determining who you are, what you're interested in and how to serve ads to you.

[0:08:52.1] JS: Yes, exactly. They all sell it and share it between amongst themselves.

[0:08:56.9] JM: Yes. There's some market that is opaque to us, but I'm sure if we looked into it we could find out how sleazy it is and how much data sharing there is. Not to not to cast stones across the entire ad tech market, but there is a lot of improper behavior, I think you would agree.

[0:09:13.7] JS: Yes, I agree.

[0:09:16.1] JM: What happens if I use a browser-based ad blocker on beefrecipes.com?

[0:09:23.0] JS: Yeah, so your computer will actually download all the ads and then the ad blocker will parse through its lists and decide what's an ad and what's not. Then it'll just not show you the ad. It'll hide it from your view.

[0:09:38.6] JM: What happens instead if I am using Pi-hole, your hardware ad blocker? Let's say I have that on my network and I go to beefrecipes.com?

[0:09:48.7] JS: Yeah, so any domain that tries to load an ad will be blocked before it even leaves your network, because Pi-hole knows about it before it sends it out to the internet. You can save on your bandwidth, because you don't actually download the images, or the video, or whatever it is.

[0:10:06.1] JM: Pi-hole blocks traffic at the network level, rather than the browser. Explain it a little more detail why that is significant.

[0:10:14.9] JS: Well, the bandwidth savings are one of the greatest benefits, I think. In addition, you're not – the browser plugins, they still get your information and Pi-hole will prevent them from even getting in the first place.

[0:10:29.3] JM: If my smartphone is connected to my Wi-Fi network, rather than my cellular network and Pi-hole is also connected to that Wi-Fi network. If I understand correctly, Pi-hole can block the ads that would get served to my smartphone, right? Pi-hole has to be on your Wi-Fi network?

[0:10:51.6] JS: Yes.

[0:10:53.2] JM: Okay. Let's say I did not have Pi-hole. How this traffic flow through my Wi-Fi network and into my phone? Give a diagnosis for how that works and then we'll talk about how it differs with Pi-hole.

[0:11:07.1] JS: Yeah. Usually your router will act as the gateway. Your phone will connect to your router and then it will go out to the internet and vice versa.

[0:11:17.5] JM: You mentioned this DNS query that occurs. Explain where the DNS query, or DNS queries fit into this exchange of traffic between the broader internet and my cellphone on my Wi-Fi network.

[0:11:31.8] JS: Yes. A lot of the times, your DNS server in your house is actually your router. Your queries get sent there first, and since your router doesn't actually know where the website is, it has to send it to an upstream DNS server, which is basically what Pi-hole is acting as, but it's a DNS server on your local network. Instead of asking Google's DNS servers where a site is, you first ask your Pi-hole DNS server. Then if it doesn't know the answer, it will go ask whatever upstream server you set.

[0:12:03.9] JM: In the case without Pi-hole, I take out my phone, I enter in beefrecipes.com into my mobile browser and my query hits my router. My query for beefrecipes.com hits my router, and the router needs to look up the IP address that is associated with beefrecipes.com, and the router is going to go to an upstream DNS server to find beefrecipes.com, to find the IP address to hand back to my phone, to look up that information, is that correct?

[0:12:42.8] JS: Precisely.

[0:12:44.0] JM: Okay. Then what happens after my phone gets that DNS lookup?

[0:12:51.7] JS: Then it knows where the website is and it can load the content.

[0:12:55.4] JM: Does it have to go back through the router to find the content that maps to that IP address?

[0:13:04.1] JS: Yes.

[0:13:05.2] JM: Okay. Then, so in contrast with Pi-hole acting as the DNS server, does Pi-hole take over that entire process so that my phone is querying Pi-hole for the URL and then also querying Pi-hole to get the content associated with the IP address from that URL?

[0:13:25.9] JS: Pi-hole basically the whole process is the same. You're just asking Pi-hole first where the website is. Pi-hole knows about the ad blocking domains, so it'll say if this is an ad domain, don't do anything with it. If it's a real domain, send it upstream and do the same process you would normally do.

[0:13:47.2] JM: Don't you need to first get some details about the page before you know what are the other advertising-related domains that are on that page? At first, I need to find the IP address associated with beefrecipes.com and then that IP address will give me access to the page and then that page will have a bunch of scripts on it and then the scripts will tell me, then based on the scripts Pi-hole will be able to determine which of these two actually get IP addresses for as well, is that correct?

[0:14:20.9] JS: No. Actually when you set up Pi-hole, you down a pre-made list of known ad serving domains, and then when your computer requests it, makes a request to one of those domains it will recognize it and just not send it anywhere. It will send it into a black hole per se.

[0:14:40.7] JM: Okay. To revisit this; I go to beefrecipes.com, it makes a DNS query from my phone to Pi-hole. Pi-hole gets the IP address associated with that DNS query, and then what happens next?

[0:15:01.2] JS: Then the website will probably also be trying to load several other ad serving domains. Instead of going through your router and out to the Google's DNS servers, it'll just stop right at the Pi-hole, because it already knows that that's an ad serving domain, so it won't send it anywhere.

[0:15:19.3] JM: Okay. At what point does that first instance of the page get received by Pi-hole? At what point is the DNS – sorry, the IP address location found and the page content being received by Pi-hole so that Pi-hole can assess what's going on that page and which of the domains on that page to block?

[0:15:45.2] JS: Well, so Pi-hole doesn't actually do anything with the page. All it knows about is the domains. It doesn't know anything about the webpage, or what's on it.

[0:15:55.9] JM: Okay. I think I'm understanding now. My phone gets the page, my phone will make several other DNS queries and then Pi-hole makes all those other DNS queries to Pi-hole and then Pi-hole just blocks those DNS queries.

[0:16:12.7] JS: Yes. It's basically just filtering out the domains that are ad serving.
[0:16:17.3] JM: Okay. I think I understand. Why is this the best model? Why do you need a DNS server to effectively do this ad blocking?

[0:16:24.1] JS: Well, it's nice because it allows you to just have one device on your network and then you don't need to install software on anything. Your smart TVs and your phones that have mobile games with ads in them, those can all be blocked without having to install anything. It's like you can block ads in non-traditional places.

[SPONSOR MESSAGE]

[0:16:56.6] JM: Software Engineering Daily is brought to you by ConsenSys. Do you think blockchain technology is only used for cryptocurrency? Think again. ConsenSys develops tools and infrastructure to enable a decentralized future built on Ethereum, the most advanced blockchain development platform.

ConsenSys has hundreds of web3 developers that are building decentralized applications, focusing on world-changing ideas like creating a system for self-sovereign identity, managing supply chains, developing a more efficient electricity provider and much more.

Listeners, why continue to build the internet of today when you can build the internet of the future on the blockchain? ConsenSys is actively hiring talented software developers to help build the decentralized web.

Learn more about consensus projects and open source jobs at consensys.net/sedaily. That's C-O-N-S-E-N-S-Y-S.net/sedaily. Consensys.net/sedaily. Thanks again, ConsenSys.

[SPONSOR MESSAGE]

[0:18:12.9] JM: I could do this on just a Raspberry Pi. Pi-hole, the reason for the name Pi-hole is that I can install it on a Raspberry Pi. I find this interesting, because I thought that routers to – a Pi-hole, it's like Raspberry Pi is a $5, or $10 piece of hardware. Routers are like 40 or 50 bucks, at least, I think. Is there any special hardware going on in that $40 or $50 router?

[0:18:43.2] JS: Actually the Pi is usually more powerful than what's in your router.

[0:18:47.3] JM: Really?

[0:18:48.8] JS: Yeah, more RAM, more specs and everything.

[0:18:51.6] JM: Oh, so this is just a racket. The router mark is just a racket.

[0:18:57.6] JS: Could be.

[0:18:58.2] JM: No, are you being cagey here, or do you know?

[0:19:01.3] JS: I mean, well I guess, I've never looked at the specifications of routers in detail, but I know they usually don't have a gig of RAM like the Pi does.

[0:19:10.5] JM: What else does a router do? Could I just outsource all of the effort of a router to a Raspberry Pi?

[0:19:17.2] JS: Yeah, why not?

[0:19:18.3] JM: Really?

[0:19:19.0] JS: Yeah. Yeah, you can set up a lot of the routers, like run some Linux type firmware too. The same DNS server that Pi-hole uses is actually what's already installed on a lot of routers and called DNS mask.

[0:19:36.0] JM: Interesting. Router software is mostly open-source?

[0:19:41.1] JS: It depends. You can install open source [inaudible 0:19:43.3].

[0:19:44.5] JM: Okay. All right. Well I mean, maybe we'll have to do a separate show on routers and maybe we'll have to start our own router company.

[0:19:52.2] JS: There we go.

[0:19:53.0] JM: Why do you need dedicated hardware? Why can't I just install this on, let's say I have a desktop computer. I've got a Linux machine in my house that I just keep on all the time. Can I just install Pi-hole on that and will it do the job?

[0:20:07.2] JS: Yeah. You just have to tell it where – tell your devices where the DNS server is.

[0:20:12.3] JM: It's running that software bandwidth intensified if I just run it on my home desktop computer?

[0:20:19.5] JS: No. Pi-hole can run alongside all sorts of different software. It's really low resources.

[0:20:27.0] JM: One thing that you need is this white list of ad serving domains, right?

[0:20:33.4] JS: A blacklist.

[0:20:34.5] JM: Or blacklist. Blacklist, right, okay. How do you get the blacklist?

[0:20:38.4] JS: There's lots of sites out there that just, that's all they do is they make lists of known domains that you might want to block. Pi-hole just basically downloads all those lists and removes any duplicates and then generates one giant list of domains to block.

[0:20:57.8] JM: What do you think about the approach of black listing versus white listing domains that would be acceptable to go to?

[0:21:05.9] JS: Yeah. With Pi-hole, that's probably difficult, because you can take the approach of blocking everything and only white listing what you need, but you'd probably find that most of your sites would be broken because of all the additional domains that actually load in the background. Blacklisting known stuff is a better approach, I think.

[0:21:27.0] JM: When an ad gets blocked by Pi-hole, is there any money that changes hands on the ad tech side?

[0:21:35.1] JS: No. It prevents the ad from being downloaded or being sent. As far as they know, they're from there and they don't see that the query was ever made.

[0:21:46.1] JM: How much does application performance for my devices improve once Pi-hole is blocking the ads?

[0:21:54.9] JS: Pretty much everyone I've talked to notices a difference, because one, the ads aren't being downloaded so your computer's trying to do less. Then the Pi-hole actually caches the DNS requests. If you've been to a site before, it'll remember where it is, so it doesn't have to do the whole lookup process again.

[0:22:16.4] JM: Let's take a use case like me. My main computer is my laptop. I occasionally take it around on trips and stuff, but I would say 25 out of 30 days my laptop sits on my desk all day and is connected to some external monitors. You use it like a desktop. Would it make sense to install Pi-hole on my laptop?

[0:22:45.7] JS: You could if you wanted to. What a lot of people do is they'll set up a virtual machine, or they'll just leave their Pi connect it at home and set up a VPN to connect to it when they're on the go.

[0:22:57.9] JM: Tell me more about that set up.

[0:22:59.7] JS: Yes you'd basically have your Pi set up at home with Pi-hole installed on it and then you'd also install VPN software. Then when you're out on your cellphone on the cellular

networks, you would basically connect back to your home network and you'd get all the ad blocking benefits over LTE, or whatever.

[0:23:21.0] JM: I don't know almost anything about modems or routers. I'm having a hard time understanding why – from this conversation, I'm having a hard time understanding why we even have dedicated modem and router software and hardware. I know this is not necessarily the purview of the hardware ad blocker you've made, but can you help me understand why we have dedicated hardware for modems and routers?

[0:23:48.4] JS: Yeah. Most home routers are like a multi-use device. They act as a gateway out to the internet, they act as your in-home DNS server, and they also act as a switch for local traffic on your network. It's like three devices in one. That's just when you say home router, it's usually you'll get all three of those things.

[0:24:12.5] JM: Is there a reason why that is not in the desktop, like the desktop devices that we have? I mean, why can't they just throw all that hardware into a desktop computer?

[0:24:26.2] JS: Well, because then – I mean, you could, but you want all the other devices on your network to take care – take advantage of it too. It works better as a separate device.

[0:24:38.1] JM: Okay. Right, okay. I think data usage also improves quite dramatically with Pi-hole, so do any numbers, like how much is ad middleware taking up in terms of my data usage?

[0:24:55.4] JS: Yeah, we've seen stats anywhere from 5% to 30%.

[0:24:59.3] JM: 30%, wow.

[0:25:00.7] JS: Yeah, it's crazy.

[0:25:02.2] JM: Are there any security concerns of running Pi-hole? Is there anything that could go wrong if my DNS, my homegrown DNS server got hacked?

[0:25:11.9] JS: Yeah. I mean, anybody could see the websites, or domains that you're trying to visit, but it's like that with any DNS server you use. The same thing can happen on your router that's acting as your DNS server.

[0:25:24.7] JM: It's certainly a privacy concern, but are there any more severe security concerns?

[0:25:29.8] JS: Not that I can think of at the moment.

[0:25:32.9] JM: Can man in the middle stuff occur?

[0:25:34.7] JS: Well, if they get control of your DNS server that's basically what they would do. They try to send, instead of sending your queries nowhere, they might send them to a fake website that they set up that the criminals whoever set up.

[0:25:34.7] JM: Right. Then does security become an important aspect of the development around Pi-hole that you've – because you're running this open source project. Is security a top of mind for you?

[0:26:04.2] JS: Yes it is. Yeah, anytime anybody gives us a bug, we try to fix it and there's been a lot of helpful people that have actually contributed to the project too. A lot of times the community will come up with a fix for stuff.

[0:26:15.8] JM: What's an example of a security bug you found?

[0:26:18.3] JS: Recent one that comes to mind as we're basically getting some variables from a file, but at that same time somebody could put their own code in there and then we would basically be running it arbitrarily without knowing what it was.

[0:26:36.1] JM: Okay. Tell me about something unexpected, something else unexpected you've learned about computer networking from the Pi-hole project.

[0:26:43.5] JS: Well, I don't know, but I guess it's interesting to see how many domains are actually being queried and what stuff is calling home. There's been reports of someone buying a cheap Chinese phone, or a cellphone and all it does is just send analytics back to the mothership. You get a cheap phone, but you're just sending them all sorts of data that they can use.

[0:27:09.3] JM: Do you know how much the data, like how much that compares to if it's just a US phone? Because I mean, US phones are sending a lot of data back to the mothership. How does it compare?

[0:27:20.6] JS: Yeah. I mean, I that was just one example that stood out to me. Then there's the smart TVs. They try to snap pictures of what you're watching at what times during the day and then they send that stuff all back to the mothership.

[0:27:37.5] JM: That's scary. Smart lightbulbs too, right?

[0:27:42.4] JS: Yeah. Yeah, there was one lightbulb that was a ridiculous amount of queries, like millions and millions of them. It's like, I don't know, it just seems excessive to me but, I don't know.

[0:27:53.2] JM: How does a lightbulb even – I guess what insights does it have about the network? How could a lightbulb have a lot of analytics data? It seems like a read-only type of device.

[0:28:05.8] JS: Well sure, but think about it if you come home at a certain time every day and you turn your lights on, they know what time you came home, and you turn them off at night, they know what time you go to bed. They can send you at targeted ad anywhere from the time you're home to the time you go to bed, and they'll start to recognize patterns.

[0:28:27.6] JM: How would a lightbulb be able to get that data synced with your identity?

[0:28:33.2] JS: Yeah, that part I don't know. I mean, that's the ad networks and all the analytic people. That's something on their end that they can pinpoint it on you.

[0:28:43.9] JM: Have you found anything else unexpected, or creepy, or interesting about AD Tech from working on this project?

[0:28:51.2] JS: Just lots of other examples like those Smart TVs and the bulbs.

[0:28:55.3] JM: Give me another one. Give me some more 1984 stuff.

[0:28:59.5] JS: I can't think of anything at the moment. I mean, we basically had enough material for six iterations of a blog post about people finding out stuff that happened on their network that they weren't expecting.

[0:29:12.8] JM: Tell me your perspectives on the ethics of ad blocking.

[0:29:17.2] JS: I think it's gotten out of hand. They're just shoving ads in your face and stealing your information. Just rude if you ask me.

[0:29:30.1] JM: What about the idea that this is the cost of getting these amazing free services that we have?

[0:29:36.8] JS: Yeah. I mean, there are people that depend on it and make money off of it, but it just has gotten out of hand in my opinion.

[0:29:45.1] JM: What do you feel that the advertising industry does that's unethical?

[0:29:49.4] JS: I don't know if it's unethical, but it's just rude and annoying, like if I'm browsing a tech website, I don't want to see an ad for a swimsuit or something. It's like, they'll just show you whatever ad they think you want to see. Then it's creepy, like it happens with Amazon ads a lot. You'll go look at something on Amazon and then the next thing you know, the next page you load there's a 100 other similar product, so it's like they're obviously tracking you and building a profile about you.

[0:30:23.5] JM: Now the tradeoff is that in some ways, that's useful. If there is this panopticon of people learning information about me in order to figure out what to sell me. In some ways that's

useful. If it's just beneath the surface and they're a little more tasteful about how they serve me products and serve me ads, it's not so problematic, right? I mean, it could be quite useful. It's more about the offensiveness and the delivery medium of those ads, right?

[0:30:58.5] JS: Yes, exactly. Yeah. I mean, like I would love to find out about a new product of something that I like, or something similar. Sometimes you're reading a website and they'll pop-up and say, “Here, buy this.” Then you scroll down and then the page jumps back up, because there's another ad that has to load, and that's just annoying.

[0:31:20.9] JM: For this, I feel the same way when I go to somebody's blog and then I'm getting ready to close the page and a thing blinks into view about signing up for their newsletter. Why on earth would I sign up for your newsletter if you just did this to me?

[0:31:37.8] JS: Yeah, exactly.

[0:31:39.0] JM: At the same time, this is a totally different conversation than the privacy discussion. There are people out there who take, and many of them in in the Pi-hole community that have dedicated hours and hours and hours to this project that are just offended by the idea that there is a Borg out there that is collecting all of this information on me and it infringes on my privacy. Does it infringe on people's privacy? How do you feel about that?

[0:32:10.8] JS: Yeah. I mean, plenty of people feel it does, because they're building a profile about you and some people feel what business is it of theirs to know what products you like to look at, or whatever.

[0:32:27.1] JM: Do you think it's possible to opt out of the dragnet of information that's being gathered about us?

[0:32:34.5] JS: I think you can, but I don't know that they'll respect it.

[0:32:37.7] JM: In practice, it's not really possible.

[0:32:40.2] JS: Yeah, because for most people that information is already out there with several different companies.

[0:32:48.1] JM: Right. Have you had a lot of privacy conversations with people in the open source community, the open source Pi-hole community?

[0:32:54.9] JS: Yeah, it comes up a lot in our forums, that's why a lot of people use it.

[0:32:59.7] JM: How is that open source community arranged?

[0:33:01.8] JS: Well, we just have some forums, like we have our own forum software and people signup and discuss things. Then there's a sub-Reddit or Github issues, people bring up privacy concerns there as well?

[0:33:18.0] JM: What kinds of conversations are going on in the community right now? What are the hot-button issues?

[0:33:24.4] JS: Basically everything we've been talking about so far.

[0:33:26.8] JM: What have you learned about running an open source community?

[0:33:30.2] JS: It is challenging. Yeah, I think our community is one of the best ones, because we have power users that will go out there and answer questions every single day. Our developers are out on the forums every day too, but then there's a lot of work to actually keep the project running, which a lot of people don't see.

[SPONSOR MESSAGE]

[0:33:57.9] JM: Every team has its own software and every team has specific questions about that internal software. Stack Overflow for Teams is a private secure home for your teams' questions and answers. No more digging through stale wiki's and lost e-mails. Give your team back the time it needs to build better products.

[INTERVIEW CONTINUED]

[0:35:11.0] JM: How have you adjusted to that? I mean, personally we have this Software Engineering Daily open source project, and trying to run it is strange. I mean, it's especially strange for me, because we're a business, but also have this open source project, so it's a little bit confused. People probably are not as spiritually motivated to contribute to it because there's a business side to it. I feel like with Pi-hole, the motives are there's no business model for Pi-hole. I think people can maybe feel a little bit more spiritually compelled to contribute to it.

[0:35:51.5] JS: Yeah. I mean, Pi-hole is a business too. I mean, we have LLC, but everything we do is we put all their code out there. We put information about us out there that we feel comfortable giving.

[0:36:05.9] JM: What's your business model?

[0:36:07.7] JS: I don't really know exactly. We're trying to figure that out actually, because we're at a point where like we're spending so much time on the project that it's getting unsustainable, because we're just running off donations. There's at least one guy that works almost 80 hours a week on it, and every night after work, I come home and spend a couple hours online just answering questions and actually trying to improve the product to.

[0:36:36.0] JM: How does this person spend 80 hours a week?

[0:36:38.1] JS: That's a good question. He's amazing.

[0:36:41.4] JM: Is he retired?

[0:36:42.7] JS: No. He runs his own consulting business. Yeah, he just has – I think he just really likes working on the project, which is what all of us, like the reason all of us keep coming back is because we like it and it's fun.

[0:36:55.7] JM: Is it true that Adblock Plus, which is this super popular ad blocker, is it true that they have made a deal with some ad tech companies to allow certain ads to be displayed, because Adblock Plus gets paid?

[0:37:13.3] JS: Yes.

[0:37:14.3] JM: What do you think about that business model?

[0:37:17.9] JS: I think they're making lots of money, but I don't think it's right.

[0:37:21.2] JM: Probably not right for Pi-hole.

[0:37:23.4] JS: Right, exactly.

[0:37:25.5] JM: Is Adblock Plus open source?

[0:37:27.4] JS: Yes.

[0:37:28.6] JM: It is. How are they able to do that if it's open source?

[0:37:31.9] JS: Yeah, I don't know.

[0:37:34.5] JM: I guess, they just don't accept PRs for people who are trying to block specific ads.

[0:37:39.2] JS: Yeah.

[0:37:42.6] JM: Maybe Apple acquire you, or maybe like a Brave. Maybe you could merge with brave somehow. Have you talked to any of these companies?

[0:37:50.0] JS: No. I think a lot of the ones that gave gotten really big have gone the sellout route.

[0:37:56.0] JM: What do you mean?

[0:37:57.1] JS: Well, like Adblock Plus, they'll accept money or gifts to do something that's not blocking anymore.

[0:38:06.2] JM: You think if you were to be acquired by an Apple or a Brave there would be some – there would be a potential compromise there? Ethical compromise?

[0:38:16.4] JS: Yes.

[0:38:17.8] JM: Apple and Brave, that's at the core of their identity is privacy and not serving ads. They're not ad-based businesses, there are ads within Google, within Apple products, but they're quite tasteful. I've not really seen an Apple ad that has disgusted me.

[0:38:38.1] JS: I think we're we pose a threat to some of those larger ad blocking companies in the way they're currently doing things.

[0:38:44.5] JM: What do you mean?

[0:38:45.3] JS: Well cause we're basically denying them the money that they are currently getting.

[0:38:49.6] JM: We'll, you're talking about Adblock Plus. I'm talking about Apple or Brave.

[0:38:52.7] JS: Well just any. I'm talking about anybody in general. Like anybody that would try to acquire us would most likely be buying us to squash us out.

[0:39:01.4] JM: Oh, wow. I see what you're saying. You're saying this technology should be commodified. It shouldn't be part of a – within some walled garden. It should be something, like maybe a model could be license your software to getting bundled in with these routers that we're talking about, these are potentially overpriced routers, maybe they should be bundling in Pi-hole. Why not?

[0:39:27.7] JS: Yeah, that's actually one of our top feature requests out there right now.

[0:39:31.1] JM: That will be awesome. Have you talked to any router companies?

[0:39:34.7] JS: No, we haven't. We've been growing so much lately. We had just have been trying to keep our heads above water.

[0:39:41.3] JM: What's the growth – like give me some numbers. How much has the growth been?

[0:39:45.1] JS: Crazy. It started out with just me back in 2014, and now we have six basically full-time developers working on it.

[0:39:54.9] JM: Six full-time developers, so but they're open-source contributors. Does it cost – or do you pay them?

[0:40:01.9] JS: Yeah. We give whatever after expenses, whatever is left over. We give split amongst the developers whatever's left over after expenses.

[0:40:11.1] JM: Where do your expenses come from?

[0:40:12.2] JS: Hosting our website and our servers. We've had some legal fees with our – our name is trademarked and everything, stuff like that.

[0:40:20.5] JM: What are the server costs? I mean, actually you're just hosting – it's like a basic static website, right?

[0:40:26.8] JS: Yeah, we have a website. We also have a server for so people can send their logs to us if they need help troubleshooting their installation. Then we've got a lot of internal servers that we use for just running PI-hole. We have a internal chat room authentication servers for identity and accounts and everything. It's crazy.

[0:40:52.2] JM: If I deploy my Pi-hole to my home network, does that Pi-hole ever call home to your centralized domain system to report back to statistics?

[0:41:06.9] JS: No.

[0:41:08.7] JM: No. Nothing like that?

[0:41:09.5] JS: No. The only time it sends information is when you tell it to. When you send us here, your debug logs if you're having trouble with your installation for some reason.

[0:41:18.0] JM: Right. That's not too bandwidth intensive, but it must be happening.

[0:41:24.0] JS: It's just a text file that gets into us.

[0:41:26.2] JM: Right. It's happening often enough that this is costing you significant hosting costs?

[0:41:32.7] JS: I don't know if you'd say significant. I mean, it's a monthly cost because it's a service we just always have out there. I mean, some months our servers have been zero dollars, because we've gotten referral bonuses from DigitalOcean and stuff. Sometimes we haven't paid anything. Other months if there's no referrals that we have to do show for it.

[0:41:56.5] JM: Interesting. Maybe you've got what? Like a 100, or 200 bucks a month in server costs and then you've got legal fees, which probably amount to, I don't know, maybe you've spent like 10 or 15,000 on legal fees. It's maybe, what do you think in like between $30 and $50,000 a year to run this thing?

[0:42:18.0] JS: Probably not that much. No. We actually posted the profit and loss thing of what some of our expenses were on the blog post recently. I don't think it was that much though.

[0:42:28.8] JM: Okay. I mean, it's so interesting, because the biggest companies, the biggest highest margin companies in the tech sector, or Facebook and Google, because they sell pixels on a screen and it's a really high-margin business, because all they do is just get paid to display these pixels. Obviously, there's a lot of work that goes into it, but it's essentially fixed cost, because they pay for the code to display those ads and to build those ad networks and there's not as much variable cost. Once they get it up and running, it just makes them really good money. This is potentially a threat to it. I think I read that you can't actually block Facebook, right. That there's some aspect of Facebook that they make it impossible for you to block at the hardware level?

[0:43:21.4] JS: Yeah. They're basically self-hosting their ads, so if you tried to block it with Pi-hole, you'd block all of Facebook.

[0:43:29.3] JM: Yeah. In some ways that's – if more companies did that that – I mean, Facebook ads, I actually find them like they're a little more tasteful, right? Would you agree?

[0:43:38.5] JS: I don't use Facebook, so I wouldn't know.

[0:43:42.3] JM: Have you never used Facebook?

[0:43:43.1] JS: I have in the past, but yeah, I don't anymore.

[0:43:46.2] JM: Okay. Interesting. Okay, so we talked a little bit about the growth in the open source community, the growth of your expenses, the growth of the project. What about the growth in usage? Do you have numbers for that?

[0:43:59.4] JS: Yeah. Quite a while ago we were at like – our closest estimate was 60,000 installations. Recently it was when that Bloomberg article came out, we looked again and it was closer to 140,000.

[0:44:13.4] JM: Wow, okay.

[0:44:14.1] JS: That's just all public information you can get from Github.

[0:44:18.4] JM: You more than doubled after that Bloomberg article published.

[0:44:22.4] JS: Well that was before the article published. Yeah, and I'm sure it's more than that now. It's not something like we regularly check. It just piqued our interest for the article.

[0:44:32.9] JM: It seems to me that the real barrier to wider adoption is the installation process, because installing it – if installing it were as easy as installing Adblock Plus, we would have a lot more ads being blocked. Do you have any ideas for how you could ease the installation of it?

[0:44:53.2] JS: Yeah. We've tried to make it as easy as possible. I mean, you can basically run one command to install it, but you have to be somewhat familiar with Linux. Usually you have to know a little bit about your network to get it set up, so I don't know. Probably the easiest way would be to somehow have it pretty bundled somewhere with a router, or something and it would just work right out of the box. That would be probably the most ideal situation.

[0:45:21.8] JM: I guess, and this is one of the problems of building a business around open source software is if they wanted to do that, if a router company wanted to do that, they wouldn't have to pay you necessarily, right?

[0:45:31.6] JS: No, but they couldn’t use our name either, because it's trademarked.

[0:45:35.8] JM: Oh. It would just have to be ad blocking router.

[0:45:39.7] JS: Yeah. They also have to distribute our source code as part of the terms of the license agreement. Whatever code they did distribute with them, whatever modifications they made, or whatever would have to be disclosed.

[0:45:51.6] JM: What are your other potential business models?

[0:45:54.6] JS: We've actually been trying to figure that out recently. I mean, right now we're we've been trying to run just a fundraiser to see if we could make enough to live off of it for a year, or something. We're trying to raise like a $100,000 just off donations. I mean, money has been coming in, but it's not enough for anybody to live off of right now. We're just still trying to figure that out.

[0:46:19.0] JM: What is required to build Pi-hole? What are the different pieces of software stack that go into Pi-hole?

[0:46:25.1] JS: Yeah. There's basically the DNS server, which is DNS mask. Then we also have a web server, so you can login and look at the stats and everything. Then that's really all it is. I mean, at its core it's a DNS server with a blacklist.

[0:46:41.7] JM: If it's that simple, what is the 80-hour a week engineer doing on it?

[0:46:48.5] JS: Well, so he's been helping build up our infrastructure so we can keep working on Pi-hole, because all our developers are just scattered throughout the world. We've never actually met in real life. We're just working on it over the internet. He's been setting up different utilities for us to use and bug trackers and collaboration tools and whatnot. Then actually helping out the users takes a lot of time too.

We put a pretty big emphasis on that, because our community is really what makes us who we are. Without them, we won't really have a product. We try to help them out as much as we can in the forums. Then after all that's done, then we can actually start working on some of the code, fix bugs or add new features.

[0:47:41.2] JM: What are some of the bugs that are being fixed and some of the new features that are being developed?

[0:47:45.8] JS: Well we usually fix the bugs as soon as we can, so there's not a ton out there, and the ones that are out there are pretty minor. Then we try to go off feature requests that people have out there on our forums. People vote on them, so we can see which ones people are clamoring for the most.

[0:48:07.4] JM: If the product is so simple, what are some of those newer features that are being focused on?

[0:48:13.5] JS: Well, one of them was the installing it on the router. I don't know, there's so many out there I don't even remember what they all are.

[0:48:21.7] JM: Okay. Well, what else is in the future for the Pi-hole project?

[0:48:25.2] JS: Right now we're working on FTL DNS, which is basically version 4 of our software. We basically took DNS mask and forked it and we're just adding our own Pi-hole special sauce on top of it.

[0:48:39.8] JM: Like what? What kind of special sauce do you have to add?

[0:48:42.2] JS: We've made it, so you can increase the cache size, so you can have a little bit more speed. Then we're also adding some guides out there on how to add your own DNS resolver, which might sound confusing. You have a DNS server in your DNS server, but we're basically setting it up, setting up a guide so you can instead of asking Google where your website is, you can just ask Pi-hole and it will be able to go and find it for you. That way, you don't have to send any of your information to anybody except yourself.

[0:49:18.9] JM: Can you contrast that? Why would you have to be asking Google and how would that contrast with – what would happen if you had that Pi-hole special sauce?

[0:49:30.4] JS: Basically, the only DNS servers out there that know about where a website is the Microsoft knows where all of its servers are. Google knows where all of its servers are. Facebook knows where all of its servers are. You have to basically go out and ask a public server where those are. It's really confusing. I don't know how well I'll be able to explain it.

[0:49:54.9] JM: Well, I think what you're saying is that by virtue of the DNS lookup process today, you'd have to touch these companies that are incentivized to be tracking you. You're

giving them information about what you're looking for just by virtue of looking up a single webpage.

[0:50:15.2] JS: Yeah, that's a good way of putting it.

[0:50:17.7] JM: Yeah. That seems like a lot of work to be able to disintermediate that.

[0:50:22.2] JS: Yeah. I mean, DNS is a – it's very complex, so it's nice to be able to have Pi-hole to be able to install it and get these benefits.

[0:50:32.9] JM: Well, Jacob I want to thank you for coming on Software Engineering Daily and for starting Pi-hole. It's been really great talking to you.

[0:50:38.1] JS: Yeah. Thank you so much.

[END OF INTERVIEW]

[0:50:42.7] JM: If you are building a product for software engineers, or you are hiring software engineers, Software Engineering Daily is accepting sponsorships for 2018. Send me an e-mail jeff@softwareengineeringdaily.com if you're interested.

With 23,000 people listening Monday through Friday and the content being fairly selective for a technical listener, Software Engineering Daily is a great way to reach top engineers. I know that the listeners of Software Engineering Daily are great engineers, because I talk to them all the time. I hear from CTOs, CEOs, Directors of engineering who listen to the show regularly. I also hear about many newer, hungry software engineers who are looking to level up quickly and prove themselves.

To find out more about sponsoring the show, you can send me an e-mail or tell your marketing director to send me an e-mail jeff@softwareengineeringdaily.com. If you're a listener to the show, thank you so much for supporting it through your audienceship. That is quite enough, but if you're interested in taking your support of the show to the next level, then look at sponsoring the show through your company.

Send me an e-mail at jeff@softwareengineeringdaily.com. Thank you.

[END]