How Google Tackles IT Security and What You Can Learn from It

Sponsored by Google Apps

Speakers: Eran Feigenbaum, Director, Google Apps Security
John “Four” Flynn, Lead of Security Monitoring, Google Inc.
Eric Sachs, Product Manager, Google Inc.
Brad Taylor, Gmail Spam Czar
Serena Satyasai, Product Marketing Manager, Google Inc.

Moderator: James Hilliard

James Hilliard: Hello, everyone and welcome to TechRepublic and welcome to our webcast today, “How Google Tackles IT Security and What You Can Learn from It.” This webcast is brought to you by Google. My name is James Hilliard, your moderator for today’s event. I don’t think it’s any surprise that today’s corporate computing environment really demands that IT pay a lot of attention to a wide range of security issues, and Google no exception. Google’s security team has to constantly evaluate a lot of the new trends and technologies that are around to protect the Search, Ads and Apps infrastructure. This knowledge and expertise really protects the data and information of, as you all know, tens of millions of consumers and business apps users, and G-mail, and Google docs, also in Google Checkout, just to name a few of those offerings. What we’ve done today is assembled several members of Google’s security team to talk about some of the emerging trends and best practices, and how Google implements them, and how you can benefit from them in your IT environment.

I want to introduce our panel. It’s a big one today. We’ve got their names and titles on the player for you. It’s also on the screen right now. Joining us is Eran Feigenbaum. Eran is the Director of Google Apps Security, and Eran, we really do appreciate you taking time to joining us today.

Eran Feigenbaum: Thanks! My pleasure to be here!

James Hilliard: Also on board is Eric Sachs. Eric is joining us and is part of the team, been a good 15 years or so, focusing on things dealing with user identity and security, and hosted web applications, and Eric, I really do appreciate you being on board today as well!

Eric Sachs: Absolutely! Looking forward to the webinar!

James Hilliard: Also here is John Flynn. We’re going to call him Four. That’s what he goes by. John is the lead for the Google Security Monitoring group there, and John, Four if you will, I appreciate you being on board and joining us as well here!

John “Four” Flynn: Thanks James! Hi everybody! Glad to be here!

James Hilliard: And, we wrap it up with Brad Taylor. He is on board. He’s got a cool title. He’s G-mail’s Spam Czar, and Brad, thanks for taking time, breaking out of your day and coming on board for this webcast today!

Brad Taylor: Well, thanks for having me!

James Hilliard: Absolutely! What were going to be doing folks is really going through a full event today. We’ve got a lot of content that we’re going to be sharing with you, definitely taking the full hour here. What we’ll be doing is talking a little bit upfront here about the security of Google Apps. We’ll talk about some of the underlying technologies, ways to authenticate. We’re going to talk about some of those modern security monitoring techniques. Four is going to take us through that material. And, then we’re going to spend a little bit of time there at the end with Brad talking about G-mail spam filtering, and then get to some questions. We’re probably not going to get to a whole ton of questions today, but what we do have planned is we’re going to keep all of the questions that come in, wrap them up in a report, get them back to this panel and their team members, and get some e-mail responses out after the event. But, I really do encourage questions to come in from the audience. Get them in early and often. Sean has already submitted a few and we’ve seen a couple of others coming in here. So, appreciate you all doing that. Keep them coming into the presentation.

A couple of things you can do on the player, ask those questions using the “Ask a Question” box, lower portion of the player. Just make sure you click that blue “Ask” bubble. A couple of ways to get access to the slides as well, click the “Download Slides” button or click the "Enlarge Slide” button, if you just want a larger view on your player today. You will need your popup blocker turned off to take advantage of that. Several related resources, I’m going to be talking about those later on in the presentation, but you are encouraged to click on those to get some more information, and those are found on the right-hand side of your player.

I also want to introduce today Serena Satyasai. She is back with us, no stranger to TechRepublic. She is a Product Marketing Manager with Google, and what we want to do with Serena here is really just take a few minutes and set us up, kind of start us off. If there is anybody that is not familiar with some of the offerings from Google, Serena is going to get us set and move them forward there. So, Serena, with that, I want to welcome you as well. And, I’m having a little issue trying to push out your first slide. You should be able to, if you just click that “Next” button. That should move us right on to the Google Solutions for IT, and we can get going with our content.

Serena Satyasai: Great James! Thanks for being here, and I’m really excited to hear from our security expert today. As many people on the webcast may know, Google Apps is one of the offerings from Google, who are actually part of a larger team that provides many solutions for IT, including our Google Search appliance so that you can use Google.com behind your firewall to search data across multiple repositories. We offer business versions for Google Maps and Earth that people can use for store locators and other kinds of information, geographic services, and of course Postini, which was an acquisition that Google made a couple of years ago, of a leading spam fighting solution in the cloud. It also offers a cloud-based archiving service as well.

Just a quick note for some of the folks who may not be familiar on what Google Apps actually includes, is that it is a suite of messaging and collaboration applications that are available through Google Apps Premier Edition. G-mail, of course, is the best known of our messaging applications, and we'll hear a lot more about its built-in spam protection, from Brad later in the webcast. But, the messaging suite does include integrated calendar and instant messaging, and our collaboration suite includes Google Sites, which is a really nice way to create team spaces with easy web publishing keys, and also Google Docs, which are online versions that enable real-time collaboration, for example, and word processing, spreadsheets, and presentations, and then Google Video for business, which is sort of like having You Tube, but for your domain. And, all of these applications are based on our cloud computing platform, and of course with Google Apps Premier Edition we provide customer support, we provide a 99.9% service level agreement, and some other features as well, such as features built in by Postini for content monitoring and filtering, and policy-enforced TLS.

With Google Apps, we serve a variety of customers. Whether you're a small business or a large enterprise, you will benefit from the same suite of applications, and as well the security topics that we're talking about today. We also have, for our EDU customers, a Google Apps Education Edition that is actually free and has been deployed to thousands and thousands of students across some very large universities, so we encourage you to take a look at that.

And, with that, I will hand it over to Eran.

Eran Feigenbaum: Thanks Serena! Today, companies have choices where they get to store their data, and the typical two-case scenario, it's in their company servers, which they manage, they upgrade, they maintain, or in the cloud. And, we have different flavors of that, and outsourcing, and letting somebody else manage your environment for you, etc., but those are the two main flavors. One of the questions that comes around a lot is about the security of the cloud. Is the cloud more or less secure than the traditional environment? I think personally the cloud is more, secure or can be more secure, than most organizations' traditional servers, and today you're going to hear a little bit, at a high level from me, about how we do security, as well we're going to dive into some very specific topics about that. But, really when you think about cloud computing and security, it's changing the mind. It's a mindset change in the way we do business, very similar to the banking paradigm where hundreds of years ago we used to hide our money and our jewelry underneath the mattresses, and then when we came home we could lift the mattress up and look and make sure that our jewelry was still there, that our money was still there. With the advent of banks and safe deposit boxes, and ATMs, etc., people got more and more comfortable with giving their valuables, their money, their jewelry, to the banks, and all they would get back is a piece of paper. But, the banks had the economies of scale. They had the armed guards, the safes, the specialized security systems, etc., that we at our home could not really afford. The same is true with cloud computing. Because of the multitenantness and the shared of the environment, cloud providers can offer, and as you'll hear today from some of these experts, do a lot more about security than most typical organizations can allow themselves to do on their own, just because of the economy.

I spend a fair amount of my time talking to CIOs and CISOs all around the world and I ask them how they feel they're doing with security, and which areas do they think are tough for them and they're behind the times. And, I hear two consistent themes. The first theme is the data problem. Security is really tough because users want to access their data from anytime, anywhere, and because of that, in the traditional environment, they're taking their data with them. They're putting their data on their laptops, on their desktop PCs. In fact, 60% of corporate data is sitting now on unprotected laptops and PCs. According to the FBI, one out of every ten laptops is lost or stolen within the first year since it was purchased. So, it's not so much the loss of the physical asset but it's the loss of the data, both from the availability perspective, "Now I don't have it," but also who may have seen it. When I was the Chief Security Officer of a major financial services company, I told my staff when making security decisions make it easy for users to do the right thing, and they tend to. If we give them the appropriate technologies and make those technologies easy to do, they will. Another way users are taking their data with them is putting it on USB keys. They're convenient. It's easy to take with me. I can work from home. I can work while I'm on vacation, etc. But, 66% of us admit to losing those small, convenient USB keys, with 60% of those having corporate private data on them. By putting the data in the cloud, I don't need to take that data with me. I don't need to store it on my laptop, put it onto a USB key. I can access it anytime, anywhere, from any internet connected device. As a matter of fact, the presentation that you're seeing today was created in the cloud. I modified it on a different PC, never storing it on another PC, and then we collaborated on it while it was in the cloud, and none of us stored it on a single PC.

The second reason chief security officers tell me they spend so much money on security and security is so tough is the patching problem. We all know major software vendors issue security patches, some on a regular calendar basis, and some on a little bit more ad hoc basis. According to the statistics, it takes companies between 25 to 60 days to deploy a patch after it's been released. Most of the CIOs and CISOs I talk to say they wish they were that good, that for major OS or application patches it takes them closer to three to six months. Now, that's time that a known vulnerability exists in your environment. That vulnerability that's being exposed and trying to be exposed by all of the hackers, crackers, tigers, lions, and bears, not to mention about all of the money that companies are spending on just the maintenance of security because of the patching. This problem goes completely away when you move into the cloud, because you have no more servers to patch. So, it's very important to understand, obviously, that you're not just giving this problem to your cloud provider and now it becomes their responsibility, but also understand how they are doing that.

So, at a high level, how does Google look at security? So, we take a look at security from a people, process, and technology perspective. First of all, we hire some of the world's foremost security experts in the field, people that have come up with innovative research around malware and browser security. Our people have a—all of our staff go through security training and have a robust code of conducts that we publish externally. If you go look on our website, on the Investment page, you'll see our Google Code of Conduct that all of our staff have to adhere to. But, from a process perspective, we were a company that was born in the internet, in the internet era, with our technology and processes built with security as part of the core DNA, as opposed to trying to retrofit security as a last moment item. So, a lot of the technologies that organizations use today did not have security designed from the beginning. Right? Most of the operations, our applications, our operating systems that we use today, security was built later. Even the internet, security was not designed upfront. So, we've taken a very different approach to that and put in the security from the front end, having a secure code development process to ensure that our products are secure and stay secure, that we write good code, and not only us looking at our products and the security of them, but also having external auditors, both from a SAS 70 as well as external penetration testing.

But, a key element to our security is the technology and the way we use our technology. We custom build our hardware. I believe it was Gartner that said Google is the fourth largest server manufacturer in the world, but we build our own hardware, just for our own purposes, which allows us to have a very homogeneous environment and not have the problems that we talked about, in terms of patching. Right? Part of the reason companies spend so much time in trying to understand if a patch is going to break something, if it's relevant for them, is because of the heterogeneity that they have in their environment, different flavors of operating systems, different flavors of applications. By having one standard gold image, that makes that problem a lot easier. The second is the redundancy that we have, built in from the software level all the way up, having data replicated into multiple data centers and not having a single point of failure, either dependent on a single data center or a single drive or rack. By having the redundancy and replicating data into primary and secondary locations, and primary and secondary drives, we can have that type of redundancy in place, as well as making sure we tightly control our network perimeter, making sure we have a good understanding of our ingress and egress points into the internet are all essential in having a good security stance.

So, people always ask me, "Why should we trust Google and why should we trust your data with Google," which is a really good question. Hopefully, today you're going to get some understanding of a couple of key areas that we decided to highlight that clients always ask us about, but security and the privacy, confidentiality, integrity, and availability of our clients' data is paramount to our business. Right? We use the same environment that we store our clients, our consumers, our business paying customers, our business free customers, are all stored in the same environment as we store our own data, and manage that environment the same way. We have a strong history of advocating for user privacy and a strong privacy policy to back that up. That's also available online for anybody to review. I know we have a lot of great stuff to talk about here today, so with that I'll pass it on.

James Hilliard: Sure Eran. I just want to jump in here and let the audience know what I plan to do. Because we've got this broken into several different segments and a lot of different speakers, I'm going to try to mix in a couple of questions between each speaker area here. I think that will allow us to get some of this content in here. So, Eran, with you, I want to start off with one of the first questions that came in, from Drew, wondering which way do you recommend authenticating into Google Apps, getting a little technical here with Drew's question.

Eran Feigenbaum: That's a great question, and Eric is going to talk about SAML authentication and OpenID, and Oauth, which we support. What we see most customers looking at, at least in the enterprise space, is comparing SAML versus our native authentication, and there are some advantages and disadvantages to both. Right? So, if you are doing SAML authentication, obviously you need to stand up some kind of environment that supports authenticating those users. But, it also allows you to add bolt-on type security elements and extends the security of Google Apps, for example to add two-factor authentication. So, we have clients using certificates and onetime passwords, as well as announced several offerings in our marketplace with some partners that offer easy to implement SAML implementations that directly authenticate users into Apps. Obviously, by using SAML as opposed to using the native authentication, you're adding another element that you need to manage and understand, and so it's something organizations need to consider. There is no silver bullet on the right decisions and hopefully you'll get more insight on that from Eric's conversation, but—

James Hilliard: Cool.

Eran Feigenbaum: —Drew, that was a good question!

James Hilliard: Let’s get Eric Sachs on board right now, and again, Drew, I do appreciate your question. Others, keep those questions coming in. We’ve got a lot of good ones already in the queue here, from Sanjay, Keith, Paul, Peter, Jim, many others, so keep those questions and comments rolling in here. And, with that, Eric, again welcome back in and let me move to the first piece of content for your portion of the presentation here, and I’ll turn it over to you.

Eric Sachs: Great! Thanks for that introduction! So, I’m going to talk about a few different security protocols related to user login, and here we have this fictional company, Shoes ‘R Us. For enterprises like Shoes ‘R Us that leverage SAS vendors, one common question, as was just asked, is how to manage their employee login, and this becomes an even more complex question if the enterprise uses multiple SAS vendors. The default option is you go around manually creating accounts for each employee at each SAS vendor. But, what frequently happens is that administrators forget to remove the access of ex-employees, including the ones that were fired and might be unhappy, or the administrators just forget to enforce good policies about changing passwords. So, the standard industry technique for this problem is the setup of a federated login scheme, to use that SAML protocol that’s been mentioned, where the enterprise runs a single login system and each SAS provider redirects employees to that login system to be authenticated. Historically, really only large enterprises have run such a federated login system and only really the large SAS providers supported them. What we’ve seen over the last few years is many medium sized businesses have tried to set up their own federated login service by say deploying some software on a single Windows server in their office. But, unfortunately, what frequently happens is that the CEO tries to log into a SAS vendor while he was on the road and that single federated login system is down or just inaccessible, even though the SAS vendor systems are fine, and this usually only has to happen once or twice before that system is removed. So, what are you supposed to do then, if you’re not a large enterprise? Well, there is another option which is growing quickly in popularity that I want to talk about. There are now vendors who will operate a federated login system as an internet service, which they run in a highly reliable manner, vendors like Ping, TriCipher, Simplified and others provide these services today. As more and more companies deploy these federated login systems, it also is having the benefit that it’s leading to more and more SAS vendors accepting these federated login systems.

Now, one of the challenges is that the security standards in this space, such as SAML, allow enough customization that some additional configuration is still required for each additional target SAS provider. So, the SAML vendors I mentioned amassed a lot of time in learning about those specific configurations for each SAS provider so that you, as CIOs or enterprise IT administrators, don’t have to. Now, Google has also become very active in these internet standards groups, such as OpenID and SAML, however our goal is really focused on enabling a more standardized and automated configuration process. For the last few years, we have accepted federated logins from larger enterprises who outsource their e-mail to us. However, more recently, we have started to do the reverse, which is that we’ve started to run a basic federated login system on behalf of the smaller enterprises whose e-mail we host. So, as an example, some enterprise SAS vendors, such as Zoho, Manymoon, Social Walk, and others now provide a Google Apps login button on their website, as shown here, and enterprise users simply click that button, get asked to enter their domain name, and is redirected to Google Apps to be authenticated via the OpenID protocol. They are then immediately signed into Manymoon or another enterprise SAS vendor without having to do the traditional occasion of username password at that website. Now, unlike the other SAML vendors I mentioned, we do not support the same level of customization for each SAS vendor, but are instead just promoting the standards of the space that we expect over time to accelerate adoption of federated login, by making it easier to set up and deploy.

So, to summarize, there are really four primary options for federated login. The first, of course, is you may decide that, given the size of your company or other factors, that it's just too early for your company to try to deploy this. However, if you are a large company, you may decide that you can handle purchasing software from a SAML vendor and running it reliably, but in terms of reliability shoot for the 409s and up that some of the SAS vendors provide. But, the other option that we just discussed is if you are a smaller business you might outsource to one of these identity vendors that I mentioned, who provides a service offering. Or, if you already use Google Apps, you can leverage our federated login service. In fact, you might even combine options three and four to get integration with the largest number of enterprise SAS vendors.

Now, another thing that Eran alluded to earlier is that one of the advantages of running your own login system is that enterprises can experiment with stronger forms of authentication than just passwords. Even though some SAS vendors, like Google, have very advanced systems to protect against brute force password guessing, it is still possible for hackers to steal a user's password from other places or use phishing techniques. In the last year, we've seen a significant increase in the number of enterprises who are deploying two factors of authentication, such as a password and a token that the user carries with them, which displays onetime codes. To be honest, there are a lot of usability barriers to getting those tokens deployed and getting employees to use them, however the vendor offerings in this space are significantly improving, and so recently enterprises are seeing much higher acceptance rates of these techniques by their employees. So, as an example, the company VeriSign has created software for a large set of different mobile phone platforms that can generate onetime tokens so that users do not need to carry a separate device with them, and this helps because users are much less likely to lose their phone than say another device that they have to keep in their pocket.

Now, if you just do decide to go down that route of two factors of authentication, there is a significant, important issue to be aware of. If your SAS vendors support software apps installed on a PC or mobile phone, then it's quite common for those apps to be hard coded to ask for a username and password. One of the most common examples would be e-mail clients, like Outlook, which use the POP IMAP protocol. Now, one workaround for this problem is to have complex machine generated passwords that employees must use from those installed software apps, and because of their complexity users are somewhat less likely to be phished for these passwords. Well, another workaround that is specific to the BlackBerry is to use REM's Enterprise Server, which can identify the specific phone making a request, as well as the person who owns that phone. As an example, that is how the Google Apps connector works with the BlackBerry enterprise server. Another advanced solution that some SAS providers are starting to use is to have the software apps actually launch a web browser as part of the login as well, and then that flow can redirect your central federated login system, which can ask for information beyond a password, such as a code from one of these onetime token generators. In Google's case, we are experimenting recently with adding such an option to the Google Apps Sync software we provide for Microsoft Outlook.

So, that brings us to the end of my section of today's webinar. I hope this information was helpful to you. And, if you want to learn more, there are some hyperlinks that are shown in the webinar. But, now let me turn it back to our moderator to see if there are any questions I can answer.

James Hilliard: Yeah, Eric, a couple in here, one actually going back to the title of our event, How Google Tackles IT Security And What You Can Learn From It. Joseph wants to know does Google use two factors of authentication for its employees.

Eric Sachs: Sure. So, Joseph, we do use it for some remote access to our network, especially VPN, however we do not use it for accessing our e-mail systems. As noted earlier, there have historically been usability problems with its adoption in an enterprise, and Google places a very strong emphasis on usability. However, we are very heavily involved in the industry work that is addressing those usability issues and, as I previously mentioned, we are updating our own installed software to better support stronger forms of authentication.

James Hilliard: Michelle is online with us, Eric, and wants to know, and kind of lays out this scenario that says that her company runs its own central login system but also syncs passwords to Google. Is that something that's okay? Is it encouraged? Am I risking any security issues here?

Eric Sachs: So, we do see a lot of companies doing that. They get a lot of the usability advantages of single sign-on to taking that approach, and by syncing passwords you avoid the problems with a software like Outlook that has to send a password. However, if you take that approach, it's also very important to make sure whatever tool is being used to sync to the Google provision API also automatically deletes accounts for e- employees, the fired employee situation that I mentioned at the very beginning.

James Hilliard: Alright. We're going to move on here. We're just over halfway through today's presentation, still more content to come. I want to bring on John Flynn, Four, is with us. What he spends a lot of his time doing is monitoring security, and so we're going to turn things over to him for a few minutes here and really talk about some of the most modern techniques that are available, and with that, Four, I want to turn the floor to you.

John "Four" Flynn: Yeah. Hi! Thanks James! Hi everyone! Thanks for joining us today! One of the things that people don't tend to talk a lot about, I've noticed in security, is security monitoring, and I think it's important. In fact, I've spent about six or seven years of my career focusing strictly on this problem, first in academia and now in the enterprise space, working here at Google and managing our security monitoring efforts here. So, I wanted to give some—try to impart some things I've learned along the way about security monitoring and how one might go about doing this in your own enterprise, some lessons learned, if you will.

So, one of the things, of course, that you want to first think about when you want to think about how to do security monitoring is what kinds of threats do you face, where do these threats come from, and how best can you design a monitoring system to face those sorts of threats, to identify them as it were. So, clearly, the traditional way of approaching this is the perimeter approach, where you sort of identify ways into your network from the outside world and try to focus your monitoring efforts there. Another approach that people consider a lot is the insider threat. "Do I have somebody within my company that, for some reason, might attempt exenterate data that we care very much about?" Right? And, then of course the third one, or a third one at least, is one that I'm sure you guys all know very well, and that is the problem of malware and the client side exploitation. So, clearly malware, in particular, is a big problem that we've seen cropping up lately and over the last few years it's become more and more prevalent, and more and more difficult to detect. So, anyway, those are three of the things that we tend to focus on, and I'll talk through some of our approaches to identifying those things here.

So, the question becomes security monitoring—I think it's important to consider security monitoring as a subset of the general security program that you have at your enterprise. A lot of times, I've come across companies that say, "Well, you know, we bought this vendor's IDS and we stuck it in front of our servers. I think we're good." And, of course, that's not true. I imagine most of you are aware of that. But, the key piece here is that security monitoring works hand-in-hand with prevention. Obviously, a good vulnerability management lifecycle, a good patch management lifecycle, are pretty key and you don't want to think of security monitoring as independent of those things. Additionally, one of the things I've learned over the years is that you can automate a whole lot of the monitoring problem, more so than I think a lot of people can conceive of, but nonetheless, at the end of the day, I think there is no substitute for having a couple of very smart—not just a couple but a team of very smart people looking at the output of your monitoring system, and we'll go into a little bit about what that monitoring system might look like in the next slide.

So, monitoring and incident response is another key thing to consider. Right? So, those of you that have had, in the past, to deal with incidents in your enterprise, heaven forbid of course, it's important to consider the lessons that you learned from incident response can be rolled back into your monitoring program. Right? So, it's important to have a low friction way of having those teams collaborate. And, finally, one of the things that I think is really important is people ask, "Hey, how much online data should I have for monitoring historical things in my enterprise," and as a rule of thumb I like to say around three months if you can, if not more. And, the reason I say this is because those of you that are familiar with the Verizon Data Breach Report might know that—at least, the companies that they've surveyed—and that's an amazing report. I recommend you guys read it. Of the companies they've surveyed, the typical time in which a company detects a compromise is actually three months after the fact, which is unfortunate, but it's an interesting and important data point to be aware of.

Okay, so let's talk a little bit about how one might build an effective security monitoring program. So, the first thing, of course, is to determine what set of data to collect. Right? And, there is a whole slew of these types of data, which I won't get into complete detail here, but there is host-based data, there is network-based data, there is log data, and so on, that you'd want to consider collecting. Right? So, it's important to kind of frame things in terms of the threats that we talked about earlier and what kind of data that you'd want to collect, based on those threats that you've modeled. Further, I think the key piece here is creating a collection and aggregation system for the raw data, and this is, of course, easier than it sounds but you want to make a way of putting all of this data in one place. So, make that a Syslog architecture in your environment. Those things are very important, is collecting that data in a single place.

Okay, so those first steps are fairly straightforward, but the tricky parts come after that. So, what do you do once you have all of this data in one place? Right? Well, the way we think of it is that you would create a set of analysis agents, as it were, that run across this data to look for signs of misuse or of compromise in your environment. But, the other thing that's kind of an interesting way to think about this is that you can sit in a room and brainstorm with your team a whole lot of ways of doing this, and you'll come up with a pretty good set of ways to look at your corpus of data for signs of compromise. But, the other thing that's really nice to have is a way of browsing this dataset to sort of discover new ways of indicating compromise. Right? So, it's not just enough to sit in an empty room and come up with ideas on your own, on how to do this, but also to provide a mechanism by which you can interact with your data and discover ways that might be indications of suspicious or otherwise anomalies in your environment.

Okay, finally, I think an important point, too, is to have an aggregation system that ranks the output of these analyzers so that SAM or SIM is oftentimes what people use for these things, but essentially to have a way that you can interact with, to look at the output of these agents that are running against your corpuses of data, to look for these signs of suspicious activity. And, then one of the features of such a system, of course, is not just to provide a series of alerts, which is actually not super helpful on its own, but also build it in such a way that it provides enough context for the human analyst to understand how that alert sits in the environment, you know, is this normal, is this not normal, and so on.

Okay, so I want to talk about this next slide here, as a lot of the types of things I see from people out there that are struggling with a solid security monitoring program. I think one of the key things is having a system of identity. So, what does that mean? It means having a system of inventory, for example, of IP address mappings. One of the things you'll find oftentimes is a network-based analysis system for security monitoring as its primary identity attribute and IP address. And, so having an inventory system that the organization depends on to be up-to-date is really important. So, this is probably one of the most important points of this talk. If you come up with an inventory system that everybody else is supposed to use within your security team and you try to force it on everybody else, nobody is really going to keep it up-to-date, and thus it's not going to be very useful for your needs. If you, however, come up with a system or use a system that's already in place, that there is an organization impetus to maintain, in other words it provides benefit to the operations teams, the other IT folks in your organization, that is when those inventory systems will actually become reliable and thus useful for your monitoring program.

Alright, quickly then, I'll run through the rest of these. Scaling, of course, is a big problem. Sometimes, if you're a medium to a large size enterprise, you have a large corpus of data. Definitely, the cloud is the way to go here. We benefit from a lot of the infrastructure that Google has built, in order to build some of these amazing apps that we've been talking about today. Things like MapReduce is something Google came up with in the first place. Hadoop is the open source version of that. Those are the kinds of things we're talking about to do scaling. And, then environmental noises, I talked about this goes hand-in-hand with prevention. The better handle you get on your prevention and your patch management, and so on, and controlling the kinds of crazy things that your users might be doing, makes the detection problem easier because the noise level that's reduced.

Okay, finally, security monitoring trends. I'll just talk a quick bit about that. Some of the things that we've seen, I've personally seen, over the last couple of years, I think the traditional approach of doing intrusion detection using signatures on the network is something that is still useful but is becoming less so. I mean, I think a great example of this is about a month ago or so we heard about, publically, the first piece of malware that uses Twitter as its command and control point. The more and more as we go along, the malware and knowledgeable attackers are going to be using very benign things to hide themselves within the network traffic, and so network analysis, I think, is going to trend towards being of limited use over time, and we're not there yet today though, of course. Antivirus, I think, is not quite enough yet on endpoints, or no longer enough on endpoints. The sort of signature-based approach is something that seems to be, unfortunately, losing. It's an important part of your overall strategy, but I think you—my recommendation is to consider approaches beyond simply antivirus as your endpoint security mechanism.

Some other things I'll just note really fast is that I've seen some cool work recently on ways of qualifying alerts. You know, the big problem with intrusion detection and security monitoring traditionally is false positives, and so having a virtual machine, for example, is one interesting way of doing this, where you take some traffic that looks suspicious, you run it through a virtual machine that's identical to your users' machines, and then you can gain a lot more insight as to what that traffic was actually doing to that machine, that it changed the machine state and so on, and I think that's kind of a cool approach. And, I think that's it for my section here.

James Hilliard: Four, Jules is on the line with us and wants to see if you can give a couple of more examples of some data that Jules should be collecting for monitoring.

John "Four" Flynn: Oh, well that's a good question! Of course, this is going to be dependent somewhat on your environment. Right? So, some of the types of things that I think are really important are things like NetFlow. You know, I did say that network analysis is trending downward, but NetFlow still is a really important data set. Obviously, you want to have that as a least amount of sample rate as possible, ideally not sampled, although that's a hard problem. Things like DNS transactions, a lot of times people don't think about that, but in your network if you could somehow have a corpus of DNS transactions historically, you can use that to do things like analysis, fast flux DNS analysis of query patterns. That's a great one. Obviously, authentication, history of log data of other types, besides just authentication data, and then various host-based logs or otherwise various host-based data I think is the way to go there. That's a couple of things there. Thanks, Jules, for that question!

James Hilliard: Right, and Four, thanks for the answer there! We're going to move on here and bring on G-mail's Spam Czar, Brad Taylor, and Brad we're already getting questions in the queue here about some antispam efforts, etc., so what I want to do is turn things over to you for about probably 10 minutes here and we'll talk about some of the spam fighting efforts you guys have going on there at Google, within G-mail, and then we'll have a few minutes for some questions at the very end.

Brad Taylor: Okay. Thank you! So, I was just thinking get the slide here. Alright, so a little bit of background on G-mail. G-mail has been around a few years now. We have tens of millions of G-mail users. We get at least a billion messages every day, so we have an incredible volume of mail to spam filter, which is actually a blessing, because from this data we're able to learn a lot about trends in spam and figure out how to filter it properly. And, what we hear from people is that our spam filtering is one of the main reasons why people like G-mail, so we think we're doing a pretty good job.

Now, I want to talk today about one of the keys behind this spam filtering is e-mail authentication. It may surprise people to hear this, but it boils down to something very simple. If you can just find out—a domain sends you mail. If it actually came from that domain, it makes the spam filtering job much easier, and you can look at it as the reverse, which is if you know what the good domains are, then everything else becomes suspicious.

So, there are two main technologies for this. There is one called DKIM, which came from an earlier effort called DomainKeys and some of them called SPF, and we were—G-mail was an early adopter of both of these technologies. So, the way that DomainKeys and DKIM work is that the sending domain cryptographically signs the message and then the recipient domain can look up, in DNS, the public key of the sending domain and see whether the signature matches. So, SPF is another way of doing this, which is the domain publishes in DNS, instead of publishing a public key, publishes IP addresses that are allowed to send mail on behalf of that domain, and then the recipient domain can simply say, “Well, which IP connected to me when it sent this message,” can look in DNS and say, “Oh, that happens to match the one that the domain published, and so therefore it’s a valid e-mail from that domain.”

Now, given this information, once we know that a domain truly sent the mail, we can do reputation on it and we can say, “Well, this domain appears to send mostly spam, and therefore it goes into the spam folder,” or, “This domain appears to send mostly non-spam, and therefore it’s probably not spam.” And, you could also detect forgeries very easily as well.

So, one of the things that people like to talk about is should I do DKIM or do SPF, and really you should do both. They have their advantages—they each have their advantages and disadvantages. SPF is really easy to implement, so I think there is no reason why someone shouldn’t do an SPF. Sometimes, I hear from people, it’s like, “Oh, you know, we’re working on a plan to do SPF.” It’s like, “Just do it.” I mean, it’s just it’s been around a few years. It’s tried and true. It’s not perfect, but it is definitely a tool that G-mail and many other places can use to identify whether your mail is valid or not. So, in this graphic here, I don’t know if you can see it, but in the graphic I show the case of eBay sending mail to G-mail. In this case, DKIM will pass because we can verify the signature, and SPF will pass because the IP did indeed come from eBay’s IP.

The second case is when the mail is forwarded. So, eBay—you signed up for eBay as a Yahoo user and then later on you decide, ”Well, I want to use G-mail instead of Yahoo,” so you forward your Yahoo mail to G-mail. So, now when G-mail is getting this message from EBay, that was forwarded through a Yahoo mail account, the DKIM will pass because they’ll check the signature and the body of the message hasn’t been changed by Yahoo, so it still cryptographically verifies, but the IP that connected to G-mail was not eBay’s, and so it will fail SPF. In that case, as far as G-mail is concerned, it’s still a valid message from eBay. So, the DKIM thing is actually pretty critical because a lot of people are migrating to G-mail, and so we have a lot of people with forwarded mail, particularly that important protocol.

So, the easiest way to explain how authentication benefits you is the anti-phishing case. So, we, by policy, we have talked with eBay and PayPal and they’ve told us—they said, “Anything that claims to come from eBay or PayPal that doesn’t authenticate, we want G-mail to not even accept.” They’re not even talking about sticking it in the spam folder. It’s just like don’t accept the mail at all. So, the only mail that can claim to come from PayPal and eBay, that comes into G-mail, is mail that actually came from those domains, and either pass SPF or DKIM.

So, this is a small step to like restoring trust in e-mail. When e-mail was invented, 30 years ago or whatever, they never thought it would be abused like it is today. I mean, I think had they known that, I think they would have designed authentication on day one, and that's what we're doing now, is we're slowly doing that. It's very difficult to upgrade the internet, but it's coming along. So, there are many other financial institutions that are very interested in doing what eBay and PayPal have done to protect their brand, so I expect to see some stuff happening in the future with this. And, the last thing I want to point out is if you have G-mail, there is a G-mail Lab feature that you can turn on that will show you this little authentication key, and what that means is that by policy with eBay and PayPal, or whoever else comes on board later, we've agreed to drop all authenticated mail from that domain and only accept their valid mail, and so the key is telling you this is an extra secure message that you're getting.

So, I want to talk about spam now, because it's easy to explain the phishing case. It's a little harder to explain the spam case, and people may not understand how these authentication protocols are useful for spam as well. So, in G-mail, our definition of spam is what our users say it is, and so our whole job is to optimize it so that users click "Report Spam" less, or alternatively the "Not Spam" button in the spam folder. We just want to put the mail in the right place for them. So, given that, all we have to do is determine the reputation of the domain. So, if it's a domain that sends mail to G-mail users and they don't tend to report spam on it, it'll have a good reputation. Or, if it's a domain that sends mail to G-mail users and they always put it in the spam folder when they get it, well it's a bad reputation.

So, there are other cases as well. So, if it goes into the spam folder automatically by G-mail and the user says, "No, that's not spam," then that means that we made a mistake and we'll correct it. So, there are sort of +1s if G-mail put it in the right place and -1s if we made a mistake. And, from that, we can compute a reputation score. So, 100% means you're a perfectly good domain that never sends spam, and if you're that good you can actually get whitelisted by G-mail automatically. We don't have to maintain white lists ourselves. It just sort of all happens automatically. And, a 0% would be a very poor reputation, in which case it goes straight to the spam folder. So, sometimes the question comes up, "Well, can't spammers abuse authentications?" Yeah, well we want them to use authentication, and then their mail goes straight to the spam folders, exactly what we want. So, domains that don't authenticate at all become very suspicious.

So, yeah, just to describe more in detail what I said before, is we have a couple of thresholds. We have a good threshold and we have a bad threshold. So, a domain like PayPal.com, which doesn't really send too much spam, if any, it goes above our good threshold and is in that green area, and so when we see mail from PayPal, we can say, "Well, it's automatically whitelisted. It's a good reputation and it goes into your inbox." And, then alternatively, a domain like SpamPharmacy.example.com, aren't too many of those because, like I said, spammers don't like to identify themselves, will go below our bad threshold and we can say, "Well, that's spam," and so straight to the spam folder. And, then there are some domains that are in between. So, something like Hotmail, which is an open signup domain, it has a lot of good users but it has some spammers as well, and so it's not enough to say, "Well, the mail actually came from Hotmail." We have to dig deeper into the message and use other techniques to determine whether it's spam or not.

So, that’s it for what I have to talk about. I have a couple of references here about DKIM and SPF, but there are references for the other things as well. So, now I’ll take it back to the moderator now.

James Hilliard: Yeah, Brad, now I’ll put up on the screen, right now, some, again, additional information. Here’s the deal. On the right-hand side of the player are the related resources. They are all clickable, and it’s actually in iFrame because you’ll want to scroll through there. There are about eight or nine links that we have in the player, so you can continue to look through. What you can also do, if you click the download slides player on this “Download Slides” link on the player, you’re going to be able to get a PDF of today’s presentation. Included in that PDF will be this slide, that has again some of these additional links, so you can get over to Google Apps, you can get over to the official G-mail blog. We also have the Google Enterprise blog there as well, and again a few of these other sites, to really get some more information, so I encourage you to do that.

A few minutes left. I want to get into a couple of questions here. Four, I want to bring you back in. Mary Ann heard you say that antivirus on the endpoints, it’s not enough, so she wants to know what is.

Brad Taylor: Yeah, thanks! That’s a pretty good question. And, again it’s one of those things that’s really going to depend on your enterprise situation. Right? So, obviously the first order of business is to add in strong preventative controls, making sure that you have a tightened down group policy, if you will, if you’re using a Windows enterprise, you have control over who has administrator access on your machines, for example, you have a pretty good lifecycle, with regards to patching, all of those sorts of preventative controls, you turn off unneeded services and do the normal system hardening type of stuff.

But, thinking about antivirus for just a second, let’s talk a little bit about that, as a detection system. Antivirus is something that really started off as a signature-based system that detects things that are known bads. Right? And, essentially you can kind of think of it as a blacklisting system that says, “These are the types of things I know about that are not so good.” Obviously, things have gotten more sophisticated in antivirus, in the ensuing years, with more modern antivirus systems allowing for heuristic detections, which means that they look for a set of patterns that are done by certain executables, and memory accesses to certain parts of the kernel and system calls, and so on. However, in practice the heuristics are somewhat noisy, in that it’s difficult to determine if something is really bad or if it’s something that is just matching a known pattern, but is actually not bad. So, heuristics, again, are sort of a mixed bag. Some promising—I’ll just drop an idea of one promising idea. If you think of antivirus traditionally as being a blacklisting approach, some people are turning towards a whitelisting approach, and again this is dependent on if it works in your enterprise. But, if you can say, “These are the set of things that I allow to run in my environment,” that’s a little bit easier problem, if you think about it, than trying to say the list of things that you don’t allow to run. So, that’s one possible thing to consider.

James Hilliard: Right. Excellent Four! Appreciate it! Hey, Eran, I want to get you back in here. We’ve heard a couple of examples of the types of organizations, verticals working with Google. Peter wants to know about any governments. Are they customers?

Eran Feigenbaum: Yeah, so we have a lot of different verticals, pretty much spanning the whole vector from education to financials, to healthcare, as well as government. Probably one of the most famous government clients that we have is the government of the city of Washington D.C., but also we're starting to work more with federal agencies, with Apps. We announced, a couple of weeks ago, that we are part of the Government Cloud Initiative, which is the government moving towards the cloud, as well as we announced that we are in the process of getting our FISMA certification, for those people in the government should know what that is, and we'll have that certification and accreditation package done by the end of this calendar year.

James Hilliard: Excellent! I appreciate that, Eran! On the screen right now, again a few more additional links to get some more information, as well you can sign up for a free 30-day trial for Google Apps. So, again, those links are all clickable on the right-hand side of the player. Use the iFrame to scroll through, to get access to all of those. Also, we're going to be sending a reminder e-mail out to everybody after the webcast. That will bring you back to the on-demand version of today's presentation. So, if you showed up late, if there is something that you didn't really clearly hear and you want to review, you'll be able to go to that on-demand version starting tomorrow. We also encourage you to send this along and get some other folks involved. If you have a colleague that you think can benefit from the event, click the "E-mail Friend" link on your player right now. That will send them an invite and they'll be able to come back to the on-demand version.

Brad, I want to come back to you. Andre is online with us today. "G-mail antispam is great! Postini antispam is also great! But, we as customers need to have a unified antispam. Is this in the roadmap?"

Brad Taylor: Yeah, it definitely is and we are working on better integration between Postini spam filter and G-mail spam filter, so yeah, that's definitely a known issue that we're working on and you should, at some point in the future, have a better much better time with them both.

James Hilliard: Alright. Excellent! Hey, we had a ton of folks, well over 400 people on the event here today with us, and that means also a ton of questions that we just aren't going to get to. We're wrapping things up right now. So, what we're going to do, get a report on over to this whole Google team and the panel. We'll try to address these questions and get some e-mail responses out to our users, but we really do appreciate you submitting all of the questions and taking time to join us for today's event. Again, look for that reminder e-mail to get back to the recorded version. Share this with colleagues that you think can benefit from today's presentation. For Eran, and Eric, and for Brad and Serena as well, the entire Google team, we all want to take a quick moment and thank you for tuning into the event today. We will keep the player open for a couple of moments so you can continue to click on some of these related resources. Also, you can keep on dropping in some questions and comments to us. Again, anything that comes in today we'll wrap up in a report and get on over to the Google team. So, again, thank you all for joining us. My name is James Hilliard, and we will talk to you down the road.