NETEVENTS

APAC PRESS & ANALYST SUMMIT

First Draft

Artificial? Or Intelligent?
Where's the Emphasis in Today's AI Security Space?

In the Hot Seat: Bryan Gale

Vice President Product Marketing, Cylance

Interviewed by: Manek Dubash

Editorial Director, NetEvents

Thanks, Manek. Alright, so as Manek said, my name's Bryan Gale. I run product marketing, field marketing globally for Cylance. What I'm going to show you briefly before we go into the hot-seat session is really a very simple example of why we believe there's a new approach needed for solving the endpoint security problem in the world at large.

If I use one of the points that Christian talked about earlier in the day, and it was really about the need for change and change really being driven by panic or fear inside of an organisation. So Cylance was really founded by a lot of folks from really the traditional AV industry that got tired and frustrated of trying to change the industry from the inside. You've got too many incumbents that are resistant to change, they're resistant to trying new innovative approaches, and the problem is they're fundamentally failing at protecting their customers.

So I'll use a couple of stats that are easily identifiable, easy to look up. So while everyone was sleeping last night, how many new variants of malware do you think were registered with avtest.com? AV-TEST is an industry third party testing house. Any guesses? Just last night, the last 24 hours. Yeah? A thousand. You're starting small. No. So in between those two. Three hundred and ninety thousand. That's just within 24 hours.

Now these are threats that they test and validate and ensure are actually malicious pieces of code. Now the bad guys don't rest so 356 days a year they're constantly churning out new variants. So that's about 140 million samples just in a given year.

Now I would posit that that number's actually a little bit low in the sense that these are only the samples that people know about. These are only samples that people have actually seen in issuance out of an enterprise, inside of a consumer desktop and they've gotten those samples then uploaded to somebody that can actually test and validate those for sure.

So I'm going to do a simple demo here that actually is going to show and prove why it's nearly impossible for humans and signature AV to keep up with that threat landscape with it ever evolving. So the set-up that we have here, I've got four individual virtual machines that are running on this laptop. You can clearly see that three of them are running a traditional AV product that uses signatures, and if you are familiar with how signatures work, just like a human virus, signatures require that a virus be found detected first: there's always a patient zero, there's always a system that's gotten impacted first before a signature can be created. Then that signature has to get pushed out to every single endpoint for that given company in the form of a DAT file.

Now at Cylance we come at it from a different perspective, a different approach. We do not use signatures; we don't use a DAT file; we don't require system scanning. Everything we do is based on machine learning and artificial intelligence. We use Amazon's cloud, a massive compute cluster, to essentially train our machine learning models, and then we effectively miniaturise those down onto the endpoint in the form of a very, very lightweight client. It uses about 1 per cent CPU, about 35 megs or so of memory.

Now the test that I'm going to run, I'm going to execute, literally run just under 100 samples of malware. These are live viruses that we downloaded this morning so they're very, very fresh. All of the competing products up here have full DAT content, they have a full network connection; you can't quite see it here but they are connected to the internet so they're able to use their phone home capability to try and query to see if anybody knows anything about the hash of the file.

I've actually handicapped myself where I don't have a network connection on my virtual machine up top. And since I'm using again machine learning and artificial intelligence, this is a client that's months old at this point so it hasn't been touched, and I'm running offline and I'm going to be executing 100 samples, just under 100 samples, of malware downloaded today. So there's physically no way that I could know anything about these files a couple of months ago when I compiled and built the agent that's running there.

So you see a simple command prompt pops up, and what you ideally want to see is an access denied here. I know it's kind of hard to see on the small screen like this but you can see that most of the others are allowing these to run completely unfettered. So the problem is once you allow malware to execute and to run, no matter what kind of solution you have, if it's behavioural or if you're watching for changes on the system

and then you want to try an attempt to roll back changes once you deem that something malicious has happened, the problem is there's too many threats out there today that that rollback capability just is not physically possible.

You look at the case of Crypto or ransomware variants where after that malware runs the entire box has been totally encrypted with an encryption key that's necessary to decrypt it that you don't have access to unless you pay the attacker in bitcoin or whatever currency is necessary.

So we see some - quite a bit is happening here. If you look at the task managers there, again the resolution is not great, but you see that a number of processes have actually been allowed to run. That's really indicative of the fact that everyone is doing a hash check or a look-up on that value on that particular string. They don't have access to it and so they allow it to run.

So another interesting facts and figures, again something that's easily verifiable, latest Verizon data breach investigation report. The number of unique malware samples to an organisation across all of the attacks that they analysed in that report, 70-90 per cent of the malware was unique to the organisation it targeted.

Now what that means is that the malware that's attacking my organisation is fundamentally different and unique from the malware that's going to attack every single one of your organisations. Now malware uses a technique called polymorphism. So that's really the ability of a virus or the fact that it's simple or trivial for attackers to create a new payload for every organisation they go after, which means that every organisation it hits, it is unique to that organisation, which means that the fundamental solutions that rely on a hash look-up is simply going to fail. They will have never seen that threat before; even though it may be the exact same virus, the hash value is different across both of them.

So you can see okay something's happening with McAfee there. We do have a ransomware sample in this sample set. Since we download these on a regular basis, every 24 hours, we don't necessarily know what kind of mix of malware we're going to get so we got lucky here with this one. And you can see that not only did it allow most every single one of the samples to run and execute but one of those samples was an actual Crypto variant, which means this box is now no longer usable.

So if we look at the fine print on there, since the resolution's not very good, typically it's some sort of fashion hey, all of your files have been encrypted, you need to pay the attackers in the form of bitcoin; there's usually a specific amount that's dictated, there's an account that you're given access to, to deposit the bitcoin into. Then if you're lucky you get the decryption keys to be able to get your files back.

Not always the case. We've started seeing some examples now where attackers are actually going back and saying okay, well thanks for paying us but that wasn't enough money; we now need x amount more before we really give you your decryption keys. Too many other vendors in the space as well are literally throwing in the towel on protection and they're just saying it's a really hard problem to solve and so you better have good back-ups and that's really your only solution to this type of problem.

Now Crypto and ransomware variants isn't necessarily as prominent in Asia Pac as it is in North America, as it is in EMEA, but I would articulate that it is very, very shortly going to be coming to this region and the globe, and it's not something that anyone should really shy away from.

So that's it for the quick demo. I'm going to have Manek come up and I guess we'll go in the hot seat now and so let's fire away from questions within the audience. Anything's pretty much fair game.

Manek Dubash

Okay. We're going to have a couple of chairs brought up so while the chairs are being brought up we can do a little sing and dance. Let me ask you the first question. I said in the introduction that your slide talked about the AI, the artificial intelligence security space. Tell me more about that. What do you mean by artificial intelligence exactly? Is it some kind of expert system look-up table, that kind of thing? What is it?

Bryan Gale

Yeah not necessarily a look-up table per se but AI in simplest terms is really the concept of a machine showing or exhibiting human logical behaviour in terms of being able to make decisions on its own. That's really how AI is embodied in this security solution. You've got an agent on the endpoint that's months old and it's effectively dealing with threats from today that we had no prior knowledge of two, three months ago. So it's making that collective decision, totally disconnected from the internet - we work great in an air gap environment - and so everything is self-contained within that endpoint agent being able to make that determination as to whether something is malicious or not and should not be allowed to run.

Manek Dubash

So I've been in this industry quite a while, as have a large number of people in this audience, as I can tell from the greying hair, and I…

Bryan Gale

Or no hair.

Manek Dubash

Yeah same here. I've seen a lot of snake oil in that time. So you're talking about a system that is lightweight, it doesn't need an internet connection, it uses artificial intelligence and still works better than any other system.

Bryan Gale

Sounds too good to be true.

Manek Dubash

It sounds too good to be true.

Bryan Gale

Yeah. So honestly our first comment to people is do not trust us. Treat us as just any other vendor. Test for yourselves within your own organisation. Don't believe our claims at face value. Now we happen to feel that we've got a better approach that works better than anything else that's out there to date. We are fundamentally the only solution in the endpoint space that's utilising machine learning and artificial intelligence solely as the means of detection and protection on the endpoint itself.

So we absolutely urge customers or prospects to try it, test it on their own. We do a number of tours that we call the Unbelievable Tour. We do these across the world at this point and we actually invite people to bring in USB sticks with live malware. If you think we're baking the test, bring in a USB stick with live malware samples that have gotten by your current protection technologies and we'll test them live on stage.

Manek Dubash

Yeah but if I'm a CIO of a large organisation I want to know what's going on inside that file. What is going on inside that file?

Bryan Gale

In terms of our solution itself?

Manek Dubash

What are you doing, yeah.

Bryan Gale

So the way the technology works, again it's very, very lightweight and we run in a pre-execution environment on the endpoint. That means before something is allowed to run - this is any executable on the system. Before something is allowed to run we extract what we call features from that file, we call it the DNA characteristics of the file. Similar to human DNA we can very, very quickly analyse that through mathematical algorithms that we use and train, again back in the cloud, and we get essentially a score for that file that deems the degree to which we think it's malicious or not. Then that score dictates whether it's allowed to run or whether it's malicious and should be quarantined. We can then send all of that information back to the administration console, it's a cloud-based solution, and that information then is available for researchers to understand why we made a determination that something was malicious.

Manek Dubash

So it's something like the spam filter on my e-mail, kind of Bayesian score type thing?

Bryan Gale

Yes, except we don't use Bayesian mathematics for it. It's a number of different features. You can think of a feature being something as simple as was a file or was a

binary digitally signed, if it was digitally signed, by whom. The point at which it's attempting to load and run within memory. Is it going to do something like a stack pivot in terms of when it's running in memory? Each one of those little attributes can be added into a linear vector if you will that we can then very, very quickly compute a number of different algorithms on it to compute that score.

Manek Dubash

What about false positives? That's a big problem with security software.

Bryan Gale

Yep. So our false positive rates within production are definitely lower than the industry norm. We're something like point four zeros and then a number in terms of the actual legitimate false positive rates across our entire installed base. It is possible for the machine learning to make a mistake; there's a lot of very valid third party software and products out there that do interesting things. They may do interesting things, and by interesting I mean they may have poor coding characteristics. They may do things in memory that are similar to what malware may attempt to do. If they generate their own buffer overflow for instance, that's a common vulnerability mechanism that's utilised.

So there are very legitimate solutions out there - very legitimate products out there that potentially we could FP on and we have a very unique approach in terms of being able to deal with those in terms of trying to whitelist them inside of the environment where we're not whitelisting them based on the digital cert that's used but based on the fingerprint that the machine learning model creates for that file, which is much more unique than something as simple as a digital certificate.

Manek Dubash

Okay. I'd like to get onto what that AI actually means in a wider context in a second but first I'd like to talk about the business model you've chosen which is to go for the enterprise first. When you consider that everyone is pretty much agreed, as indeed your colleague said earlier, that most of the malware problem is through the endpoint, so in that case why aren't you protecting most of the endpoints which is, let's face it, consumers. So why go for the enterprise first? Is it just the money?

Bryan Gale

It's definitely not the money.

Manek Dubash

It's not the money. You heard it here first.

Bryan Gale

No I think fundamentally again one of the points that Christian raised earlier in the morning was some of the most innovative companies in the world are really focused

on the why of what they're doing, not the what. For us it's really about protecting every endpoint on the planet. We just so happen to choose the enterprise first because frankly that's where all of our founders and 90 per cent of the executive team came from.

I've spent the last 16 years in the large enterprise security space at one of the big incumbents, as did most of our executive team. That is the landscape that we know, it's the problem space that we know well.

Also there's very few examples of enterprises or start-ups if you will that start out in the consumer space and are able to adequately swim upstream into the large enterprise space. So we just chose based on our heritage to start at the large enterprise and go down market from there.

Manek Dubash

Okay. I've got one final question, and then I'm happy to throw it open to the floor, I'm sure you've got lots of questions out there, which is, as I said, it's about the whole artificial intelligence thing, the implications it has for society for things like robotics, for connected cars and so on and so forth. Is there anything that you're developing - does it have a wider relevance?

Bryan Gale

Yeah it absolutely does and I'd say the field of AI and the field of machine learning has broad applicability across so many different industry sectors. It's not just about defeating the next type of board game, the Go champion that got defeated by Google's AI machine. We just happen to have taken this and applied it to the endpoint security space. The core intellectual property that we develop is around machine learning, the data science team that we have. That data science team is not a group of scientists that have come from the security landscape, they are theoretical mathematicians, statisticians, and we apply that data science to the field of endpoint security.

There's so many other areas that are definitely applicable here in terms of the way we can take this technology into other adjacent security areas that we plan on doing in the future as well.

Manek Dubash

So would that be like a spin-off?

Bryan Gale

Not necessarily a spin-off but we realise that we've got to continually stay a step ahead of the attackers. They are constantly evolving and changing their approach if you will. Whether it's something that we go after the network space, whether we go after the identity or authentication space, there's a number of other adjacent areas that we can make a very logical leap.

Manek Dubash

Thanks. Questions? There must be some. No questions? Any questions? Yes.

From the floor

Hi, I'm [unclear] and come from Network magazine, Taiwan. So I have a question. Is it possible the security competition will be trained like good AI and bad AI? As I know, the network was generated by automatic tools. So I want to know in the future is it possible?

Bryan Gale

To have good AI versus bad AI?

From the floor

Yeah, good AI like you to protect us, the bad AI like automatic tools that generate malware to attack us. So is it possible the competition will be trained to [unclear] AI?

Bryan Gale

So you say the competition, so I'm going to assume that you mean the attackers, not necessarily the security industry competition. So I think it's very plausible. The attackers are a very well financed industry by themselves. They are after financial gain in most cases. Sometimes obviously it's nation state attackers but many of the attackers at large are really after financial gain. They have well-funded development teams, well-funded QA and research teams as well and they are always looking at new ways to attack enterprise networks as well as consumer endpoints.

So I think it's certainly in the realm of possibility. I don't know that we've seen that yet but I have no reason to believe it won't happen.

From the floor

David Heath, iTWire. One of the big impediments I think for you guys is that your competitors do so much more than just anti-nasty. They've got asset management, they've got all manner of other things. So I'm guessing that you've got to compete in that space as well.

Bryan Gale

Yeah so we - if you look at any of the competitors that are out there, let's take McAfee for example. McAfee has ePO, ePolicy Orchestrators or Central Administration Console. They're going to claim their marketing tagline of security connected and they have an endpoint suite that pulls in probably 12 different disparate products. So really they're a great example of how the security industry at large has grown over time, not through innovation but acquiring individualised point products as they think that those are needed by an enterprise.

What we're seeing now is that while they've done that they've fundamentally lost their ability really to truly protect the endpoint. So we're focusing with laser-like focus on that area first. I can honestly say that we're probably never going to go out and build an encryption solution and some of the other adjacent areas are probably going to follow a similar fashion where the operating systems themselves are building in those capabilities that are more than good enough for large enterprises to be able to use. So you've got BitLocker on Windows, you've got FileVault on Mac. That's one example.

There's other adjacent areas that all of those competitors of ours have bundled into suites. Some of those things we have no intention of building, something like web content filtering, again something that we truly do not believe is needed on the endpoint. We're not going to go out and build our own host IPS solution. We're not going to try and protect against every CBE exploit that's out there because we don't care how the box has gotten exploited; we're controlling execution on the endpoint regardless of how that threat got in there.

Manek Dubash

A question at the back?

Matt Allcoat, BT Global Services

Yeah hi, it's Matt Allcoat, BT Global Services. So the machine learning, that AI thing, is really cool and we all saw Microsoft really get in a lot of trouble over this in 24 hours as their chatbot turned nasty. How is your software not going to do the same thing and how do you protect against that?

Manek Dubash

Racist security.

Bryan Gale

Yeah. How do we not let the machines unleash themselves on the world? I don't know if I've got a great answer for how we are going to prevent the chatbots from turning evil if you will, other than the way we train the mathematical models and the machine learning capabilities in the cloud, we test it for months up there. We typically will update our endpoint about every six months and so during that six-month period there's new mathematical models that are put into there; there's new information that's tested against new threats that are tested against that, and we try and validate that in a very robust fashion over the course of months before we push or essentially miniaturise that down onto the endpoint itself.

It's also the sense that the AI or the machine learning that we're using is very singularly focused in terms of preventing execution of malicious code on the endpoint. I'd say we're probably a few years away from that getting a little bit more broadly capable in terms of being able to make voice or other determinations or capabilities.

So it's through rigorous testing I guess and validation before it's pushed down to the endpoint is the best I can come up with as of today.

Manek Dubash

Question?

Naveen Bhat, Ixia

Hi, this is Naveen Bhat from Ixia. A question about AI technologies. Most of these AI technologies, whether it's neural networks or expert systems et cetera are based on machine learning from historical data or knowledge of the past. How do these apply to the security space when you're actually trying to anticipate something in the future that has not happened in the past?

Bryan Gale

So when we talk about the machine learning or training the mathematical models, all of that again I've said we've done up in Amazon's cloud. So this is a capability that probably about five years ago wouldn't really be economically viable in terms of the compute power that's necessary to do that.

You can think of what we have on the endpoint as being the collective knowledge or intelligence of every threat, every specific piece of malware in the known computing universe that we use in our training datasets to train our mathematical models.

So we have vast samples of good files that we know are good and bad files that we know are bad. Think of those as training sets that are used any time we go through and introduce a new type of mathematical algorithm or a new type of feature correlation to try and make those determinations. So it's a continual game of training those machine learning models to continually look at new files that are out there as well as comparing against the capabilities of the files that have been malicious that we have in those training sets.

So it is absolutely part of that. A big part of the training is absolutely based on historical data that we continually collect as well as new data that we continually collect. And then we physically analyse those with the models, figure out where they compute in terms of the score or result of the model output to continually update those. So it's something that we can't sit by; we constantly are having to evolve our techniques.

Manek Dubash

Okay. One more question at the back there.

Nikhil Batra, IDC

Hi, this is Nikhil from IDC. So I'd like to move a bit away from the technology to the SLAs and the contracts that you've got with your enterprise customers. So making the claims that you did can be a bit dangerous at times so what happens if a malware gets through? What were your general contracts with your enterprises are that in case a malware gets through then who bears the brunt of it and just some insights around that?

Bryan Gale

So we don't make SLAs that say we're going to protect 100 per cent within the enterprise. We just - we make the claim that we are drastically superior than the traditional signature-based approach. We are not perfect, there are threats that can still get through. What we claim is that we really raise or we introduce friction in the environment that makes it that much harder to exploit and get malicious code to run on that box.

Now in the case where something does get through, since we've removed a lot of the noise from within the system, the security ops people inside of a given organisation are no longer continually chasing ghosts within the data. One thing that we find when we go out and we get broadly deployed in an auto quarantine fashion inside of a large enterprise, their network sensors kind of go quiet because all the command and control communication that's been proxied behind some sort of - inside of their environment, all of that outbound communication goes very, very quiet. So they have an ability now to really truly focus on the singular meaningful threats in that organisation.

Again, as I said, we're not sitting by on our laurels. We're coming out with a solution within the next quarter that's really going to go after that detect and respond phase in terms of we see - we have visibility into everything that's happening on the endpoint and we're going to apply that AI and that machine learning knowledge to say that okay, something may have executed; it didn't get triggered as malicious but here's exactly what happened on the box as a result of that.

So we are going to provide those tools from a research standpoint for the organisations like that that need to go and find out exactly what's happening on their endpoints in the environment.

Manek Dubash

Okay. Thank you Bryan. I think we need to move on but before we do I'm going to conduct a little poll. Press and analysts only please. This guy is promising a new innovative approach using cool shiny technology versus the old signature-based approaches, which is why I'm not going to let the vendors vote. So all of you who think that this approach is the future and the other guys will be out of business soon, please raise your hands. One, two - not totally convinced yet.

Bryan Gale

Oh we've got three.

Manek Dubash

Three. Okay.

Bryan Gale

Test it yourselves.

Manek Dubash

There you go. Tried and tested here at NetEvents. Bryan, thank you very much.

Bryan Gale

Thank you Manek.

Manek Dubash

Okay.

[End]