aiworld Conference & Expo

TRANSCRIPT FROM THE PRESENTATION:

Autonomous Cyber Defense:

AI and the Immune System Approach

Eloy Avila
CTO – Americas,
Darktrace

I'm at Darktrace, company about five years old. I wanted to spend a bit of time talking about cyber security as a use case for machine learning and for AI and specifically why I think it's needed, why I think we'll continue to require automation and this augmentation of our human security teams, given today's attacks.

Just a few minutes on Darktrace. We were founded about five years ago out of both Cambridge with a team of leading mathematicians, Cambridge University, as well as a subset of our founders that had spent their time and their lives infiltrating and maintaining themselves in networks. So these were GCHQ, MI5, MI6 individuals that knew what the current state of the art from a defensive standpoint was and how they were easily able to evade this and maintain themselves across these networks for days, months, sometimes even years.

And so they came together with these mathematicians and said, there has to be a better way to approach the problem. And five years later we've demonstrated that it's a solvable problem, that AI has a use case in cyber security. We've got over 7,000 deployments at this point in time, over 35 offices, approaching 850 or so employees, and a fairly broad application of the technology across industries. While most of our customers, or a significant portion of our customers are in the media entertainment and as well as financial industry, a lot of that is IP and regulation. We're pretty well spread across the board both from an IT perspective as well as from an OT perspective.

And why did we get here? Why did we think a new approach was required? And I think unless we've all been avoiding the news we know that the recent slew of headlines has made it more clear than ever that the legacy approaches to cyber security just aren't quite good enough. They can't keep up. Just recently we saw another big one from Marriot. And so, why is that?

And we find that whenever we plug our AI solution into networks, whether it's small companies, or larger ones, Fortunate 500, in those larger cases, these are companies that have very large security stacks, often times tens of dozens if not hundreds of solutions already protecting them whether it's perimeter endpoint network, what have you, yet we find that in 80% of the cases, when we plug in, within weeks or months we find vulnerabilities. In many case these things have already existed on those networks for months at a time without being detected. And when we expand that to smaller companies that number raises to nearly 95%. Ninety-five% of companies have vulnerabilities already in their networks, generally undetected by legacy tools.

So we believe this is largely due to automation both on the adversarial side in that attackers are getting more automated. They're collaborating more. They're building marketplaces. They're specializing their toolkits, and then providing these as components. Some of these were leaked from government agencies, and that battle has shifted. That fourth dimension of the theater of war has shifted from the nation to nation state sort of attacks to the corporate enterprise.

These attacks are getting much higher velocity. They're working a machine speed at this point in time. They're being augmented by machine learning in some cases, and they're also getting much stealthier. Attackers are getting better at covering and knowing what those legacy tools do, whether logging information, how they're logging information and cleaning up their tracks.

The other reality is that several estimates, one by Frost & Sullivan, suggest that in a couple of years, by 2020, we'll have a fairly large shortage of security expertise across out teams. Now most large companies have three layer SOC teams, security operation center teams. They might have a layer one that does a triage, thousands, maybe tens of thousands of alerts coming into their seam or their tools, and someone's got to make a decision very quickly to say whether something's a false positive, something worth ignoring, or something that needs to be bumped up the chain to a level two or level three SOC analyst.

By 2020 we're going to have about a million and a half gap in unfilled positions from a security standpoint. So again, we believe that the speed, the voracity, the complexity, the stealthiness of these attacks warrants a new approach. The lack of shortage warrants a new approach. And we'll get into why, while tools have evolved to defend networks, I think we're starting to see some attackers now employ those same tactics to further their means.

I have a quick drink.

So we took inspiration from the human immune system. We figured as an analogy our own immune systems have a great way of telling what self is versus what a foreign entity is and giving us some signs that's something's wrong when we detect anomalies within our own immune system. So we came up with this analogy, this AI platform called the Enterprise Immune System. You know, in essence, how do we model that same approach to look at every behavior of every single device, every user, peers across that network and the network as a whole over time, establish a moving baseline, and then surface threats across the board such that someone can take action?

And that's essentially what we've done. In life and in our digital echo systems, our enterprises, attacks, careless or malicious users, compromise is inevitable. So, we approached it from the inside out assuming that you're going to get breached, assuming that the reality of our networks now is porous. It's ever-changing. The perimeter is not what it used to be. We can't just put up a wall, firewall, and expect that we're going to be safe.

Data's moving across the board from IoT devices through cloud environments, one or many, and so we wanted a solution that understood what a given network was, that specific network in a bespoke manner, and evolve over time to determine what those anomalies are, surface them, and then potentially take action. And so that's what we came up with as an approach.

Now the issue with AI and the issue with networks is that no two networks are alike. We came at it with the expectation that we weren't going to predefine what good or bad was, this whole argument between signatures and rules-based or supervised and unsupervised machine learning, if you move it toward the machine learning part of it. So we assumed from the get-go that every network should be treated as bespoke, that every network should have its own definition of what normal is, that uses and devices change over time, whether its role is centrality in the network or the behavior of that specific device changing from one mode to the next.

And so the tool has to be able to keep up with all this information. And so, the reality is, we took a couple of these different reasons and determined that AI was our course of action, that unsupervised was going to be our primary approach to it, and that we have to come at it with an assumption that we didn't want to tune everything across the board. That we didn't want to take something from your network and apply it to someone else's network, that we didn't want to take the prior learnings of someone defining through signatures what bad was.

And then the other thing, it has to scale. Networks change over time. It needs to be applicable from very small S&P enterprises, 10-person hedge funds, all the way through telecoms that have millions of devices, hundreds of gigabytes of traffic flowing through their network. And it's got to deliver value fairly quickly. Obviously AI is great. Machine learning is great. But some approaches, some models require immense amounts of data, training data, immense amounts of data sets, and a long time to train. And so we came at it with the approach that it has to deliver value nearly immediately and continue to evolve over time.

So as I mentioned, we started off with unsupervised as our approach to modeling behavior, modeling a network, modeling the devices, the users, and then how they interact and behave. And this precisely because whether there was training data or not, we wanted to ignore that side of it. Again, we want it to be applicable to every network and have it learn specifically when we plug it into that one network to the point where we shouldn't have to tell it what normal HTTP or web traffic is. It needs to figure that out itself.

Attackers are stealthy. They're intelligent. They have, as humans or even as AI's going forward, they have a way to maintain high levels of stealth, and so they can spoof protocols. They can determine what's normal on a network and then start to evolve themselves in that regard. And so we came at it with that unsupervised approach, essentially looking at raw network traffic, which is a key distinction from some other tools that might do sampling or others that might give you logs on a non-real time basis or plugging into the network or observing raw network traffic for all of those devices via one or more choke points on that network.

In real time we're extracting about 400 or so different parameters or features out of that data and then serving that up into several different classifiers that ultimately make decisions and pop-up high-fidelity alerts, anomalies for the end user for those analysts or help them make decisions.

Over those five years of being in business of serving millions of alerts, millions of events, and helping our customers with these, behind the scenes we also have a large team of analysts. These are either ex-agency or intel individuals that, again, know how to infiltrate, know how to detect some of these things. And so from these millions of events that we're witnessed, where it made sense we have applied some supervised approaches to certain models or certain ways of triaging these on behalf of our customers through the solution, but the vast majority is unsupervised.

Part of our approach also incorporates some deep learning, some multilevel neural nets, essentially, that help us make decisions across those various weak indicators or other signals coming in across those different features and then help us make a decision. Many of these attacks that are longer term stealthy, or low and slow, DNS tunneling or attacks that might take years in the making leave very small, weak traces of indicator across the board, and so we can take some of these different signals over time, over a long period of time, and keep that state and modify, attune some of those signals as well through deep learning.

"And when we expand that to smaller companies that number raises to nearly 95%. Ninety-five% of companies have vulnerabilities already in their networks, generally undetected by legacy tools."

And then the last thing we did was essentially try to build some level of trust, whether it's through exposing our decision making as part of our alerting, ultimately a SOC analyst gets an alert. They've got to very quickly determine whether to make a decision and say, "This is a false positive," or "Something's encrypting my network, encrypting my data", or, effecting my network in some potentially malicious way, and then "Take action." And so as part of our own deployment of AI we actually provide hints as that AI's making decisions to say, here's another weak indicator. Here's another one. And finally there's enough for me to say, "Pay attention to this. Take action." That's part of the initial approach.

The second thing we've considered over the last year and a half or so was we've got 7,000 deployments. We've generally proven that the anomaly detection part of it through machine learning and through AI is fairly high fidelity we're providing. Results are often times better than legacy tools that are easier to triage, easier for some humans then to make a decision in generally a lot less time than it would take to collaborate with others to look at hundreds of different log lines and make a decision.

The next logical step for us was, how do we then further augment that decision making, the action and the response to help the human teams across the board? We call it Antigena, and, again, taking that same analogy of the human immune system, our antibodies that then quickly attaché and try to respond to a threat, this is what I would describe as surgical IPS. And IPS might be a curse word in some security teams' and IT teams' nomenclature, but, in essence, if we have a high level of confidence on how that conversation should be normal, what that normal pattern of life for every communication across device is within the network, we then should be able to intercept just that one part of the conversation if we know it to be a threat or a potential threat and leave the rest of the activity behind.

Security approaches shouldn't really be about, at the first sign of threat knocking a device entirely off the network and effecting work product, effecting businesses processes, and effecting users. In fact, when we apply this approach and customers have deployed the autonomous response portion of it we often times find that users don't even know we're taking action on their devices. We get as prescriptive as a point in time duration, one port, specific protocol to another port and another IP specific protocol, stopping that conversation for 30 minutes, three hours, what have you, whether it's to give the security teams time all the way to fully integrating with your firewalls and knocking that device off the network, which is the typical legacy approach.

And some of the attacks are such that time is of the essence. Merck's network during the NotPetya and some of these other Trojans that went around and went rogue, their entire network was knocked out in two and a half minutes. It took them nearly three weeks to bring it back up, and only out of sheer luck they had one of their domain controllers still up because they had a brown outage I think in Indonesia or somewhere in Africa, if I remember correctly.

There's a great write up on Wired about it, but this is the speed at which some of these attacks are taking effect, and when you show up in the morning as an analyst and you've got thousands of these alerts and have to determine which one to look at through legacy means, there's just no way you're going to stop some of these things. And so we said, with these high-fidelity threats, if we've got a high level of confidence in detecting through AI what's happening, if we understand what the conversation is and that normal, why not take action? And in many cases through Ai we can respond, intercept those conversations, stop them dead in their tracks, oftentimes within seconds. I have a use case later today on that one.

So let me go through some of the examples. I think there was a discussion earlier about, you know, the robots taking over and how much we are comfortable letting these things start to take action. And certainly that's not different in our enterprises than in our networks. Very few of our customers say, "Yes, go ahead, turn on the AI. Turn on the robot and have the T1000 or the HAL start taking out devices and whatnot." Certainly you can imagine that being the case especially of OT or ICS scatter environments that are controlling our traffic systems, that are controlling our water purification, our power generation. And so we've got a couple different steps that we built into this autonomous response technology as part of the AI.

First and foremost we take the time to actually train it. We have to establish a baseline of what normal is underneath the network so that the AI then that's doing the autonomous response can more confidently make those choices. And so we've got a couple different modes. Essentially recommendation mode and human confirmation mode are very similar in that some customers will actually deploy it purely from a recommendation standpoint to help train their own analysts to help them determine whether they would have made those same decisions, whether they would have taken the same data attributes into account to ride out that decision, or possibly then to augment their own security policy.

In most cases they'll actually deploy in confirmation mode where they get to say yes or no when that AI says, "This is the action I want to take." Through their mobile device or I they're sitting in front of the screen in front of the solution they can then take that action and have peace of mind. And we're fairly prescriptive. Even when we turn it to active mode we can say, these types of devices, these individual users, this subnet or this subset of the network is where I want you to take action.

I will focus quickly on a few use cases that bring to mind the state of current affairs on our networks and then some sort of food for thought on the last couple of slides. This one I particularly like because it covers a lot of today's day and age as far as threats, as far as devices. You've got IoT. You've got Insider. You've got non-signature, non-malicious endpoints that would have not been detected by legacy tools.

And Intrepid Insider had access to an application with user data. This was in the medical industry. This particular individual put up a couple Raspberry Pis in the way in the ceiling tiles in the roof in the data center. Obviously, these sort of devices that are now $5 to $30, you can pay in cash, and you can find about a thousand different videos on YouTube about to use them and configure them. It essentially was a man-in-the-middle attack. They were piping application traffic through these Raspberry Pis and then pointing out to their own private server at home, essentially trying to harvest user's username and passwords.

We detected it through normal means. From our perspective it was a rare external location that was beaconing to rare connections on different ports, rare devices seen on a network. All of these things were rare to us, rare to that network, but would have been unknown and unflagged by any other tool.

Here's another one. It kind of covers some of the IoT side of it. A customer had a video conferencing device. These things have two different channels, audio and video. Someone had taken over that device. And IoT tends to be a doorway for many, many attacks to get into networks and then start to pivot laterally across the board. Many governments are using these now to get to networks. If they wanted to they'll find a way to get through the light bulb in the refrigerator and onto the rest of the network.

In this particular case the attacker appeared to be wanting to listen on to establish themselves on this, it was a boardroom in particular, during M&A activity for this company and had essentially taken the channel that was doing audio through a perfectly normal protocol, normal port, has to be open on the firewall for these things to function. But this particular device out of the others, out of its peers was behaving abnormally. And the destination for some of this data and the amount of data was, again, one that we flagged.

And then lastly an example of where we see the majority of attacks really originating nowadays, very well-crafted phishing emails in many cases augmented by machine learning tools from attackers in many cases targeting employees not just at their corporate email addresses but even their personal. And this was the case here. Someone received a phishing email to their personal email, checked it at work, inadvertently clicked on the link, which tends to happen, and downloaded a malicious executable Trojan.

Here's an example of where this customer had autonomous response, that part of the solution in place. This would have been a Trojan that would have enumerated across their files and started to encrypt, typical ransom ware, and within 33 seconds not only did we detect it, we took action and stopped it before it could spread across the rest of the network and continue to encrypt files.

Now I think I got a few more minutes. I want to spend some time going over where we think we're headed. If the last couple of slides, the IoT, the cloud, the changing environment is not enough, the lack of expertise, we are seeing the beginnings of machine learning and AI being used from an adversarial standpoint. Solutions that are better at detecting and raising that bar for the attackers as well, and there's obviously a lack of expertise even on the outside. So they're starting to apply machine learning to augment their toolsets, to authorize new forms of malware that in seconds you can go to the dark web, upload them through GAN networks, essentially modify the code, modify the signature or the hex signature of it such that it's undetectable by any other current versions of all our antivirus or endpoint detection tools.

These things are already on the black market. They essentially obfuscate the code at very low cost to whatever malware or Trojan you want to upload to that and spit it back out in a way that's going to go undetectable. We see it time and time again. And again, we believe that the only way to circumvent these attacks going forward is to rely on the network traffic as a truth up and until the network traffic itself becomes part of the attack. But in essence, no attack will be unnoticed within the network traffic over time, and it will mimic some sort of behavior that it is similar to other attacks and abnormal from that network traffic.

I'll give you an example based on some proof of concepts we've seen, and I think even IBM has come up with something like this, but jumping to air gap. We often times have different DMZ zones or separate networks. In rare cases now we actually have air gap networks Stuxnet, for example, was one example, nuclear sites, but essentially we've got AI that can listen, turn on microphones, listen to human speech, and make decisions. Payloads can be optimized based on where that malicious code wants to be, where that payload wants to be delivered itself.

And so jumping from one network to the next, it might be augmented by someone listening in or some AI listening in on the mic and determining which file to attach itself to base on what the developer's being asked to do as far as moving it to production systems, for example. We're seeing the advent of that. We're seeing the advent of that in a few other similar scenarios.

Now this is the one area where I've taken some notes. Genetic rootkits, rootkits, malware that lives underneath kernel space, undetectable by signatures, AV, other tools, but we're seeing the advent of GANs for malware generation deepfakes, which are a way of threatening cyber security, AI generation, capture bypass, for example, and I think long term my fear is that it'll be less about data exfiltration and maybe more about data manipulation that helps organizations make their decisions.

And with that I'll leave you to it. I think we've got another session later. Thank you.