Lee:	00:00:00	This episode is sponsored by Darktrace, the world's leading AI company for cyber defense and creator of autonomous response technology. From subtle insider threats to machine speed ransomware, cyber attacks will inflict more than $1 trillion in damages during this year alone wreaking havoc before security teams have time to investigate. By using artificial intelligence, Darktrace learns while on the job to distinguish friend from foe, and when it senses an attack, the AI fights back against the bad guys within two seconds. It's time to supercharge your security stack. Start a free trial at www.darktrace.com/trial.
Sean:	00:00:40	Welcome to our second episode in Ars Technica's podcast mini series on emerging uses of artificial intelligence. I'm Sean Gallagher, Ars' IT editor. This time, we'll be diving into one of the holy grails of information security, catching insider threats.
Sean:	00:00:54	There've been a number of recent high profile cases where people within organizations use their access to data for self-enrichment or ill-intent, and it slipped past the usual policies and tools that are collectively referred to as data loss prevention. Most of the time, employees are long gone before their data theft is noticed, if ever, and preventing data loss almost requires a minority report level of pre-cognition.
Sean:	00:01:18	To get some insight into how AI could play a role in detecting insider threats, Lee Hutchinson and I spoke with Kathleen Carley, Director of the Center for Computational Analysis of Social and Organizational Systems at Carnegie Mellon University.
Sean:	00:01:33	So, now, joining us is Kathleen Carley. She is the Director for the Center for Computational Analysis of Social and Organizational Systems or CASOS at Carnegie Mellon University, and she's also a professor of computer science at the university. Thanks for joining us, Kathleen.
Kathleen:	00:01:49	You're welcome. It's my pleasure to be here.
Sean:	00:01:51	So, can you tell us a little bit about what CASOS does?
Kathleen:	00:01:55	So, sure. It's a university-wide center that is focused on applying social network analysis techniques, machine learning, and agent-based modeling to solving real world social problems, everything ranging from insider threat to counter terrorism, to information diffusion.

Sean:	00:02:19	That's really interesting. So, can you tell us a bit about the insider threat research you've done?
Kathleen:	00:02:22	Insider threat, first off, is very difficult to study because, mostly, you only have data about who actually was an insider. So, we were trying to come up with this machine learning approach to be able to discriminate early on whether someone was even in the process of becoming an insider threat.
Kathleen:	00:02:44	What we found was that, in general, the psychological and financial factors, although sometimes there for different people, were not consistently the same across everyone, and that the family issues and the health issues were not consistently the same.
Kathleen:	00:03:03	The things that were consistently the same where they had to, of course, have access to the information. They had to have the opportunity to access it basically when they were alone, and they also had what we called high social betweenness. What high social betweenness is it's a network characteristic of people that says you're connecting people who otherwise would not be connected.
Kathleen:	00:03:34	In their case, it was very, very special what was happening. Imagine in your own life, you have people in who you talk to, who include members of your family. They include people that you work with. They include your friends. For most of us, some of our friends know some of the members of our family, some of our family knows some of the people we work with. Some of the people we work with know some of our friends.
Kathleen:	00:03:59	We really have this circle around us, almost like a support group, of all these interacting people. What was happening with these insiders is as they got closer and closer to the event, they started dropping connections, and they did not introduce their family members to people at work or to their friends. They did not have their friends interact with the people at work.
Kathleen:	00:04:24	So, they would maybe be connected to only one member of their family now, and only a couple of people at work, and only a couple of friends, and those people didn't even know each other. In this way, they were the only person who was connecting these people. That was critical because it allowed them to operate without any kind of oversight.
Sean:	00:04:47	So, what role does using machine learning or artificial intelligence play in being able to detect that? It sounds like

that's one of those things where you have to basically plug into factors that are outside of the normal range of data you would collect on people.

Kathleen: 00:05:04 Well, yes and no. In this case, we were fortunate because we had datasets that actually had some information on who was communicating with whom. We had their email. So, we were able to actually run machine learning on that. We augmented that data with other information on their psychological characteristics, home characteristics, health, et cetera.

Kathleen: 00:05:29 Then we basically tried out simple machine learning models where we thought we knew what the factors were, only to find they just weren't very good. Then we just ran some great big ones to discover all the different kind of variables that together made sense.

Kathleen: 00:05:46 From that, we looked at the models that resulted, and we're able to see that it was this not only this betweenness, but the change toward it, as well as the opportunity I mentioned.

Sean: 00:05:58 So, in terms of the real world implications of this, did you learn anything that could be applied to monitoring network behavior of people in a workplace that would predict their tendency to be an insider threat?

Kathleen: 00:06:15 In terms of being able to predict it, I would say that we're still far from a perfect prediction scenario because most of our data is based, like I said, we only know people who really did engage. We don't know people who almost did and then turned away, and we certainly know that a lot of people who don't ever engage might have similar characteristics. That's one of the things that needs more basic research from my prediction standpoint.

Kathleen: 00:06:42 However, while I can't say that if you start looking like this, you are definitely going to engage in insider threat. I can say that if you look like this, chances are, you're very unhappy and disgruntled, and that there might be interventions that could actually help you be healthier and happier.

Sean: 00:07:04 Okay. That's interesting. I mean there's all sorts of applications for making people healthier and happier if you can detect that they're not happy, but also, there's some level of minority report here as well, I think.

Kathleen:	00:07:18	Well, I would not use this to just go on and say they're definitely going to engage in this kind of an activity. The other thing I would say is that this actually suggested ways of organizing work, so to just make it just less likely, right?
Kathleen:	00:07:32	So, remember, the other part of this was they had to have the opportunity and be alone with the data. So, if in fact people are never alone with sensitive data, then that is itself a preventive measure.
Lee:	00:07:48	So, as you were conducting this research, did you say that you were looking at insider threat behavior that involved the exfiltration of company proprietary data for purposes of industrial espionage or did you look at that plus also more broader activities, more different insidery kinds of threats?
Kathleen:	00:08:10	So, we looked at both. We looked at both more industrial espionage as well as national espionage. We also looked at accidental insider threats.
Lee:	00:08:23	Could you elaborate a little bit? I'm curious about what an accidental insider threat means in this specific context.
Kathleen:	00:08:29	So, in this particular context, what we are looking at is individuals who had access to sensitive information, like sensitive corporate information. We're using a computer system, and may have done something like put it in Google groups or send it through an email without thinking about the fact that that wasn't a closed system. So, they forgot turn on certain passwords or something.
Sean:	00:08:58	Okay. So, that's more of a data exposure issue.
Kathleen:	00:09:01	Exactly, and there's a variety of ways of doing that. We use machine learning to model the human process of response and how humans would learn to behave in those organizations, then put that into a simulation model. So, we had lots of little machine learning models of people running around inside an agent-based simulation. Our model basically suggested that no matter how well you try to train people, these kind of accidental events were always going to happen. They were just not preventable, totally.
Sean:	00:09:36	You mentioned you looked at both industrial and national type scenarios. Is there a behavior model difference between the two types of insider threats? I mean, it seems like there's a different culture for security and the industrial space and in the

government and military space. So, I would assume that they would have different motivations as well in the cases where it was intentional disclosure.

Kathleen: 00:10:04 So, we did not build a simulation model of intentional disclosure, just of the unintentional. In terms of the machine learning models, our strongest ones were on the corporate world just because there's so much more data. So, the other one we basically use for confirmation to see to what extent it fit with the real data.

Lee: 00:10:28 There were similar results when you're looking at both industrial espionage and espionage espionage, right?

Kathleen: 00:10:36 Yeah, and we found that individual motivation was so highly varied in both settings that it wasn't in and of itself the predictive factor.

Lee: 00:10:45 That's really interesting.

Sean: 00:10:46 Yeah. Well, that makes sense. I mean, I can think of the cases I've covered such as the recent cases at the national security agency where people brought home things with them, for example. In one case, it was maybe psychologically associated, and another it was someone trying to essentially brush up at home on things so they could advance their career before they retired.

Kathleen: 00:11:13 Exactly. So, there's just a lot of motivations.

Lee: 00:11:16 So, is there a way to, I guess, is there a way to package and commoditize the methodology of insider threat detection that you specifically are working on? Is there a way for if a company wanted to begin implementing some, I don't want to use the phrase predictive analytics for insider threats, but I guess I'm going to, if a company were going to do that, do you have recommendations that they could take forward today?

Kathleen: 00:11:44 Well, one recommendation is to always work in terms of teams. Another one is that if they are going to monitor interaction, then there's a whole set of different measures they can use to easily identify when an individual is dropping interactive ties with others and starting to be too isolated.

Kathleen: 00:12:05 So, I would say look out for people who start becoming isolated. Look out for people who are not engaged in friendly team behavior. Organize yourself such that you always work in pairs,

at least. Engage in social interactions within your group, so that you build triplets of people who all know each other.

Kathleen: 00:12:31 The reason for a triplet as opposed to just having a buddy system is because if I loan you money, you may or may not pay me back, but if I loan you money and you're friends with Joe and Joe's a friend of mine also, you're definitely going to pay me back because otherwise Joe will know. So, triplets, organizing in terms of triplets is very, very important for any kind of system where you want norms to develop naturally to support your mission.

Lee: 00:13:00 I like these recommendations as someone who tries to be as privacy conscious as it's possible to be.

Sean: 00:13:07 Right. I mean, I was going to add to that that, obviously, there are some workplaces, especially like in national security environments, where you have to surrender a certain level of privacy to work with that environment, but there are obvious boundaries in the corporate world to how much is ethical to surveil your employees.

Kathleen: 00:13:32 No. Absolutely. The other thing is if we were to go completely that route, we don't know yet how fragile the machine learning tools are for doing this. I mean, we know we can trick a lot of machine learning tools very, very easily, but we don't know for the ones particularly for fighting an insider threat that we could imagine deploying. We don't know how fragile they are.

Sean: 00:13:54 Right. So, if somebody understood the rules being, "Okay. I need to maintain social networks. I need to maintain contacts with people to give the illusion of not being isolated."

Kathleen: 00:14:07 Right. Exactly.

Lee: 00:14:10 Right. Plus, I mean, again, as you stated, the algorithms that are being used, the machine learning tools may or may not be particularly fragile in ways that are, I guess, unknown.

Kathleen: 00:14:21 Correct. I also think that this is an area where a human machine teaming approach that involves really developing culture is very critical because the way in which you act as an insider now is so much more associated with the technology you use. You need to develop not only trust with the other members of your team, you'd have to develop trust with the other members of your team who aren't human. So, I think this is really one of those

areas where developing organizational norms and culture are going to be absolutely critical.

Sean: 00:14:53 Is there any particular area you think we're missing here that you would like to put out there as far as what you've seen in your research around these types of behaviors?

Kathleen: 00:15:05 So, I think there's another side that we were saying in some of the behavior that is growing different, and we definitely need to look and explore this more, and that is the recruitment angle. So, if I know that someone is vulnerable, I can now use various kinds of phishing and technology approaches to actually go and recruit people, right? So, I think the other part of this is trying to get smarter about recognizing and stopping these kind of phishing campaigns.

Lee: 00:15:40 What it sounds like you're saying dovetails very, very closely with how a foreign agent might look to recruit vulnerable employees in the United States who might be willing to give up secrets in exchange for whatever it is they might be looking for. Some foreign agents are after money, some are after a human connection to be listened to. The motivations are legion. It sounds like we're talking about the same kind of things when it comes to identifying who might be vulnerable as an insider threat and who might be co-optable.

Kathleen: 00:16:13 Yes. I think that's right. So, that's why I'm saying that another approach here that you need to do in addition is trying to help people understand when they're starting to be co-opted, try to help them understand that they might be being recruited.

Sean: 00:16:29 Yeah, and unfortunately, it seems like most of the time that we do things like that it tends to be punitive in terms of trying to sting people. I know that some of the things I've seen recently in this area as far as looking for people who are recruitable tends to be an operation to find people who are potential problems and get rid of them as employees or put them in jail. So, I guess the problem there is finding a way to do that in a way that's not a law enforcement or security-specific-driven type of process and make it more of a corporate culture process.

Kathleen: 00:17:09 Right, and more of also just part of a general educational awareness process. I think that we want to move toward security cultures, and I think that in those, a big aspect of it is improved critical thinking, improved training individuals on how do you know that you're under surveillance, how do you know you're under attack, how do you know you're being recruited, how do you know what information to trust is the other side of

watching for someone who actually says, "I don't care. I'm going to do it anyway," kind of thing.

Sean: 00:17:41 Right, because one is detection and the other is prevention.

Kathleen: 00:17:46 Right. I think that you need to do those, have those integrated together as part of your overall plan.

Lee: 00:17:53 Well, and then I was going to follow up the other side of that overall plan then is the after action thing. After you've identified an insider threat and you've spoken with the person and they've said that, "Yeah. That was me. I was going to do that," or whatever, what then is the appropriate response? Is it a punitive response? Is it a rehabilitative response? I mean, I'm sure it depends on circumstance, but overall, if the result of seeking help when you feel you're vulnerable for recruiting, if the result of that is that you are punished for that, then there's no incentive really to seek help. You're going to go ahead and do the bad thing.

Kathleen: 00:18:29 Exactly. Moreover, if you feel like you're getting punished, it makes you more vulnerable to other recruiting attempts-

Lee: 00:18:38 Yeah, definitely.

Kathleen: 00:18:40 ... from other groups. So, I would say you should treat it more as an educational thing. So, perhaps a way to think about it is to put this in the context of conspiracies. You might be like, "What?"

Kathleen: 00:18:59 So, it turns out that when people start to get engaged and start to look at a particular conspiracy theory and start moving toward it, there's actually a period where they can be helped to understand that that's really not true, that that's really just a conspiracy theory, and it's not based on facts, and so on.

Kathleen: 00:19:22 At that point, you can pull people away. One of the ways you can pull people away is you'd help them understand how things get weirder and weirder, and show them all the weird things to look for.

Kathleen: 00:19:32 On the other hand, if they're not pulled away, people would start believing in one conspiracy. You often tend to end up believing in two, three, four or many different kinds of the conspiracy theories. It's like it escalates or spirals down into an increasingly vulnerable position.

Kathleen:	00:19:48	I would say that from an insider threat perspective, similar things could ... There's a similar kind of cognitive process that could be going on.
Lee:	00:19:58	Well, it's the power of hidden knowledge. It's the feeling of being on the inside and knowing that you know something that the rest of the world doesn't know. That's I guess the appeal of buying into theories like this is it places you within the privileged ingroup of knowing these things that all of the sheep don't know or whatever.
Kathleen:	00:20:18	Right, and it's like introducing you to one recruiter, then into a different recruiter, and then to a different recruiter. It builds you a network of like-minded people who then mutually reinforce each other and mutually reinforce each other to you. So, all of a sudden, you're hearing the same thing over and over again about why you should give all this stuff.
Lee:	00:20:39	It's like we're dangerously close to slipping into a political discussion here.
Sean:	00:20:44	I was going to say this sounds very similar to the way that some of the social media information operations work in terms of trying to identify people who are vulnerable to specific types of ideas, and then using them to spread this information.
Kathleen:	00:21:03	No. It is basically. It's very similar process. As more work becomes telework, more work is online, there's not going to be this strong line between these.
Sean:	00:21:13	Right.
Lee:	00:21:15	Yeah. No. I mean, where we work, we're all virtual. I'm in my home office here, Sean's in his. We communicate a lot through keyboard, and there's the potential, I guess, for that feeling of isolation and social vulnerability to creep in there, too.
Kathleen:	00:21:31	Yes. So, now imagine you've got corporate knowledge that could be taken away. So, you start getting recruited by one individual, and then they introduce you to several others. Really, it's just a bot. You just have no way of knowing, and you have no way of knowing that all of those different others are all connected to each other.
Kathleen:	00:21:51	What I would say is create situations where you can encourage the more positive building of connections. So, encourage more days where you have group meetings, where you bring

		members of your family or where you bring in friends with you to work or encourage more group level socialization or things like that.
Sean:	00:22:17	I guess the other thing I wanted to ask was whether you saw any risk of the same patterns that you picked up in simulation and machine learning being used by an adversary through a machine learning driven system to identify candidates for recruitment.
Kathleen:	00:22:39	I think that if the adversary had the data, that yes, they could employ machine learning to identify those who were at risk for recruitment. Yes. I would say that it's not the exact same algorithms, but it would be related because the things that we'd found were both the opportunity thing and it said that the different motivations were different. What they need to be able to find is that there is a motivation, as well as the network factor, as well as the opportunity. So, that's why it's a slightly different algorithm.
Sean:	00:23:15	Well, let's hope we don't see anything like that.
Lee:	00:23:18	Well, Kathleen, I think we're just about at the end of our time here. I wanted to thank you for making the time to dial in and speak with us. We're very grateful for the opportunity and it's fascinating to hear you discuss it.
Kathleen:	00:23:30	It's my pleasure and thanks for having me.
Sean:	00:23:31	Yes. Thanks again.
Lee:	00:23:34	There's a battle happening right now for the world's most sensitive data and cyber criminals are gaining ground. They're sophisticated attacks are scanning for the slightest cracks in the digital perimeter, an employee falling for a phishing email, a cloud application left up without a firewall or even a smart refrigerator using a default password. Once they get inside, it's only a matter of minutes before your data is encrypted, stolen, or erased entirely. At this point, for most organizations, it's game over.
Lee:	00:24:01	Darktrace has changed that game for thousands of smart cities, international nonprofits, and Fortune 500 companies. With the first ever AI-powered autonomous response technology, Darktrace instantly neutralizes in progress cyber attacks that are already inside the enterprise, containing the threat without interrupting your normal workflow. Autonomous response is on

guard 24/7 on weekends and on holidays intelligently defending your data on your behalf.

Lee: 00:24:28 The reality is that the next automated attack will strike too fast for humans to amount of defense, but with Darktrace, the machine is fighting back. Find out how on darktrace.com.

Sean: 00:24:42 To dig a little deeper into whether AI can really help detect when people are about to walk off or upload their employer's data, we turned to another company that focuses on insider threats from a slightly different perspective. We talked to Rob Juncker, Senior Vice President of Research and Development at the data loss prevention software company, Code42.

Sean: 00:25:04 Rob Juncker is the Senior Vice President of Product at Code42. Thanks for joining us today, Rob.

Rob: 00:25:10 Thanks for having me, Sean.

Sean: 00:25:11 So, I read some information that you recently put out in a study on data exposure that said that in the last 18 months, half the companies that had a data breach cited employees as the cause of the breach.

Rob: 00:25:28 Yup. That's absolutely right. Recently, we came out with our independent research focused around the data exposure report, really taking a deeper look at data loss protection, if you will. What we really began to find was some interesting statistics that began to bring some light to the risks of employees and the way in which they can either cause data loss leak or theft for that matter, and the way in which those are correlated.

Rob: 00:25:51 In fact, what was interesting for us was realizing today that 79% of information security leaders believe that employees are an effective frontline of defense against data breaches, but what we really found was a lot of data that disputed that notion.

Sean: 00:26:07 So, there's been a lot in the news recently about employees walking off with data. For example, in the case of the Alphabet executive who went to Uber and also, there've been some other cases recently of insiders acting in not necessarily a manner appropriate with being a company employee such as what happened at AT&T recently. Can you talk a little bit about how that behavior emerges from the standpoint of how it can be detected? Because, obviously, it's one of those things where if you're looking to catch somebody stealing data, you don't start

looking for them stealing data the day before they leave a company.

Rob: 00:26:50 Yeah. That's a great point, and maybe just to talk high level right now and why this problem has been growing at such a dramatic rate and as we bring up that 79% number we were talking about before, there's a lot of macro changes that are happening in today's what is data economy that we have today.

Rob: 00:27:06 In all honesty as you begin thinking about the data that we all based our businesses on, that data is being weaponized, if you will, daily, whether it be in our favor or against us. Are we actually working and researching and increasing the business value that we have through data or alternatively, is that data leaving our organizations in a way where it could be used against us as well?

Rob: 00:27:25 There's some macro trends that are really amplifying the causation of this today. To begin with as you talk about why this is happening, we are seeing a revolving door that is occurring right now in the labor market as well. Right now, 40 million employees in the US quit their job, and that number has risen every year since 2010.

Rob: 00:27:46 What that means is literally at any given point, 50% of the labor force is looking for a new job according to Gallup polls that we have out there today. That means that this revolving door begins to put more and more of your data at risk. At the same time, too, what we've found is that employees today, while we give them mechanisms by which they are encouraged to use tools for their job, employees will continue to make choices that actually put that data at risk.

Rob: 00:28:12 What we actually found there was despite that people are looking at insider threat as well as thinking about how people share information rather than sticking to company-provided file sharing and collaboration tools, what we're finding is essentially one in three business decision makers are using social media platforms like Twitter, Facebook or LinkedIn to actually share information, and 43% of them are using personal email to actually send files and collaborate with their colleagues, which means that this data portability side of it is also beginning to increase, if you will, the need for people to actually be paying attention to those mechanisms.

Rob: 00:28:51 As you put it out there, too, it's important to recognize on the final side of this that the existing processes and the procedures that we have in place today failed to really stop that insider

threat. What we actually found again was 69% of information security leaders say that that data loss prevention that they're used to today cannot stop that insider threat from happening. As you begin wrapping those things up, it begins to quantify how big of a problem that you have.

Rob: 00:29:18 Now, you brought up the departing employees scenario with obviously that revolving door. That's a really big challenge for organizations to solve. What I loved, Sean, is asking organizations one very simple question, which is, "Do you have a process today to capture someone's laptop when they leave?"

Rob: 00:29:33 Typically, the answer is, "Yes."

Rob: 00:29:35 "Do you have a process to capture someone's badge when they leave?"

Rob: 00:29:38 Again, more often than not, the answer is, "Yes."

Rob: 00:29:41 Then you ask them the questions simply put is, "Do you have a way in which to make sure that you collect all of their data when an employee leaves?"

Rob: 00:29:48 The answer there is a lot of whole hums and people looking at me like a dog would look up at an alarm clock in the morning, with a, "Roop," right? That's a problem that they don't know how to solve.

Rob: 00:29:59 I think that's really where things could start coming together. Just as you again talk about those departing employees today, what we've found in looking at Code42 data is that departing employees have a tendency to actually take the information they're looking forward to their next job before they resign from their current job. In most processes to capture data today, what ends up happening is that a security team, after hearing about a resignation, will go ahead and start monitoring an employee for that data loss leaker theft. When in reality, the event where that data escapes your organization has already happened. It's before they turned in their resignation.

Rob: 00:30:37 So, for us, it's important to begin to embrace tools that look at people from how are they interacting with the data that they have and how much data are they putting at risk at any given point.

Lee: 00:30:48 I'd like to ask a clarifying question here. How many of these, when you talk about employees leaving with data, how much of

that do you believe is actually malicious, and how much of it do you believe is simply because the employee feels some sense of ownership over whatever that data is, misguided or not?

Rob: 00:31:04 Yeah, yeah. Fantastic question. I mean, as you asked that, in most cases, it's not malicious that they're actually taking the data with them, right? At the end of it, it's malicious from the sense that the company does own that information. From that sense, it's malicious. When we talk about the scope of insider threat, there are purposeful insiders that are doing breaches, if you will, that are very malicious, right?

Rob: 00:31:27 We saw at AT&T in this last year, just as an example, an insider was essentially bribed $1 million to unlock $2 million worth of phones, and essentially hack the company that way, but when you think about departing employees leaving an organization, in today's world, employees feel as though part of what they create is owned by them and they believe that it's their right to take that information from where they are today to the next role because it's part of the, if you will, intellectual property that they've assumed as they've built that together.

Rob: 00:31:58 Now, simply put legally, contractually, that's definitely not true, right? Especially as you begin to starting to talk about things which I typically refer to as core intellectual property things, for example, like code as an example for engineers. That can be highly valuable to an organization where they might be leaving with some of the most crown jewels, prize jewels, if you will, that they created while at their former employer.

Rob: 00:32:24 The thing that I bring up here and something I just want to mention is that as we talk about these mechanisms, and you talk about McAfee, for example, where insiders actually walked out the door with a lot of intellectual property on USB drives. It's not that sophisticated in terms of some of the mechanisms by which data leaves an organization, but the problem is it exists in the wild somewhere where it can be found, where it can be stolen, and there's no controls around that data at that point.

Rob: 00:32:53 So, the information that is leaked becomes highly vulnerable just as time goes on and it's unmanaged, but, legally, again, I think you have to consider it malicious in all cases because it was essentially a theft of information that belongs to the previous employer, although in some cases, the employee, it's not malicious intent, if you will, from their side of the equation.

Sean: 00:33:15 That problem gets a lot more difficult when you're dealing with the fact that you've got a lot of people who are mobile employees or maybe in their own devices, for example, or our contractors or our short-term employees, who are doing things with intellectual property that's potentially hazardous to corporate profits in longterm, but also offers other security issues since you've got a share in a different way with people inside and outside of the company. How do you deal with that sort of a threat?

Rob: 00:33:44 Well, and, by the way, you just brought up some great points in terms of how people look at employees, right? You brought up things like contractors, as an example. In a lot of cases, people don't necessarily think about what is the information life cycle process associated with contractors. They have a tendency to focus a lot more based upon the employees that exist within their organization.

Rob: 00:34:04 As you begin talking about some of the approaches by which people do that, one of the most important things that people can do is begin to think about what is an insider threat program look like, whether it be malicious or whether it be accidental, and ensure that they're getting the right level of controls, as well as process and culture in place.

Rob: 00:34:22 I'll tell you, I think that's what really begins to, as you begin thinking about that, is setting up the right culture is the very first step, right? How do you think about your employees? If you have a big work-from-home culture, if you've got a big contractor culture, make sure that you actually get insight into how those employees are working remotely, and what access, and what levels do you actually give them access to.

Rob: 00:34:47 Beyond culture, the second thing we always talk about is transparency, making sure that employees don't feel threatened by the program, but rather, make sure that they are aware that a program exists, right? The more that you aren't honest with your employees and the more that your program actually hides what's happening behind the scenes to them, the more likely they are to revolt away from it, if you will, or not embrace it, right? Convey that you're passionate about the data that they create and that you want to create it, right? Share the information with your employees and bring along those employees with that transparency.

Rob: 00:35:19 The third side of this as we talk about creating that culture as well as the insider threat program is really executive support around it, making sure that you talk about the importance of

the data that an organization creates. Then there's fundamentals, right? I love the conversation we had about intent versus malicious versus just it's their right. Fundamentals in the controls that are in place still need to be in place, right?

Rob: 00:35:46 The things like making sure you do the background checks, making sure that training and awareness is there, and as part of the entire process, not just for employees but the contractors that you bring around as well, building policy about acceptable data use thing, automating the acknowledgement of that policy so that users are required to see that. Then at the end of this, just like you said as well, is with the number of vectors that organizations are faced with right now for collaboration, and the fact that in a data economy, it's so important we enable our users without putting blocks in their way.

Rob: 00:36:22 What I mean by that is they have to have the freedom to collaborate and communicate at the speed at which business needs to operate without those walls being put up, but at the same time, you need to be conscious of all of those different vectors by which information can flow and watch them independently as well.

Rob: 00:36:40 When you begin to take all of the aggregate of information, if you will, that data corpus that someone begins to interact with and you take that corpus of information, begin to analyze it for changes, for example, if all of a sudden, and maybe just as we jump back, if you will, to some of the McAfee as an example, the finance and sales ops teams where the actual culprits in that particular case where information was leaving.

Rob: 00:37:07 If we see things, for example, where employees are downloading files that they shouldn't have access to, finance people , for example, all of a sudden touching source code or alternatively, right? CSV is showing up on salespeople's laptops, which is more indicative of a download that might be happening of contact information. You can begin to not only spot alerts that you need to take action upon, but you can begin to spot trends.

Rob: 00:37:34 Those models that we're looking at today in terms of machine learning and artificial intelligence can begin to see trends around those what I call macro sources of change, which are, what are those lead indicators that are showing things that are happening in weird ways? Literally, the data biometrics that you can create to begin to understand how users are interacting can begin to spot change before the data's taken or alternatively, as you begin identifying a departing employee, these tools like

ours can go back in time and be able to see did they take data before they actually turned in their resignation, and then take action upon that as well.

Sean: 00:38:15 How do you pick up on behaviors that say somebody may have been doing all along that may be not necessarily malicious at the time, but maybe exposing data and where maybe trickling data out slowly over time and something they've been doing all along if you come in and start with a new system to try and catch behavior that deviates?

Rob: 00:38:37 Yeah. So, as you begin to go down this path, I think one of the things I want to maybe draw attention to is it was funny in our report. What we found was that over three quarters, 78% to be exact of CSOs, and 65% of CEOs admitted to clicking on a link that they shouldn't have, showing that no level of employee is actually immune as we begin going down this process.

Rob: 00:39:00 I bring that up only because in a lot of cases, it starts with the accidental insider, right? As opposed to it being something that is, to your point, what I would call a trickle behavior, it's an event, right? In a lot of cases, tools can identify an event of bad behavior. It is an alert that fires. It's a binary thing as opposed to something that a data model needs to be looking at or analyzing over time.

Rob: 00:39:26 As you begin thinking about that landscape, today, it's impossible for you to pay attention to all of that data, and that's where all of the alerting technologies, all of the capabilities that come around with that analysis come into play in such a positive way.

Rob: 00:39:44 You ask again about machine learning and artificial intelligence. There's no simple way to represent a user's behavior anymore and how data is meant to flow, right? Information is no longer information as we used to think about data flows of the past. Just like I was referring to before, those macro indicators of change begin to enable those models to be incredibly successful in terms of how do they analyze risk and how do they spot risk sooner so that people can take action upon that, too.

Lee: 00:40:22 So, Sean brought up the cloud, which is, obviously, incredibly important in this context. I'd like to ask about another very popular buzz word that's sweeping the IT nation at this point. From an insider threat perspective, is the push out from the data center to the edge an inflammatory thing? I mean, one person's edge computing is another person's shadow IT department. It all depends on how official the push out is. I

mean, are there specific controls you'd recommend for countering the threat at the edge as that threat I guess grows?

Rob: 00:40:54 Yeah. Well, to your point, as data begins to move, we're definitely seeing traditional file servers melt away to more collaborative approaches that exist for file, sync, and share, and other activities along those lines. Here's what I would actually recommend to be honest with you is that as you think about the way in which your employees want to work. More often than not, they want a lot of tools that allow them to collaborate better, to interact more effectively.

Rob: 00:41:25 What that means is, is naturally be sure you're embracing the cloud, right? Don't fight it. The organizations that are fighting the cloud are ultimately going to be fighting their users. If you've learned one thing about the law of governing dynamics over time is that if more people are trying to break down a wall than defend it, the wall eventually comes down, right?

Rob: 00:41:43 So, you need to pay attention to what your users want. To your point, I mean, getting data to the cloud, pushing data to the edge is something that users are asking for. For what it's worth, we've definitely seen a trend where if you don't embrace it, great, they're going to use their personal Dropboxes and do things along those lines to get around it and then you're infinitely more at risk than you are if you begin to embrace controls around it.

Sean: 00:42:08 How much of an impact do you think GDPR and the recent California legislation will have in terms of how companies have to relook at their policies as far as tracking insider threats go, not just in terms of exfiltration of data, but internal misuse of data as well?

Rob: 00:42:29 I think the sweeping legislation that's out there has already caused us to think differently about data privacy, but as we talk about insider threat again and we talk about contact information and we talk about PII that's embedded in many of our sales records that we have in an organization, an extraction of that information becomes something that organizations need to pay attention to because it does risk PII or PHI being leaked out to the real world in a way in which it can't be controlled.

Rob: 00:42:58 So, the responsible organizations, they've already embraced data privacy. Organizations that are looking at data privacy are embracing it as part of data protection. I think as we begin taking a look at the insider threat, more often than not, that's where the tools and technologies will allow you to be able to

pay better attention to data privacy through that data protection lens.

Sean: 00:43:19 Yeah. Just a wrap up question looking forward, what do you see right now as an emerging area where companies need to focus attention on the potential vulnerability of their data, and to bring it back to AI once again, how does AI play a role in fixing that as far as you see going forward in the future?

Rob: 00:43:41 As you begin talking about the next generation of technologies, which really is the current generation of technologies around artificial intelligence and things along those lines, those are only going to serve to amplify the signal through the noise. As we begin to watch the way in which users interact as opposed to dictate it, we're giving users freedom and it's necessary for us to understand how they're utilizing that freedom and ensuring that they're using it in positive ways and not being the accidental insider threat, too, where someone is clicking on a bad link or any of those other behaviors.

Rob: 00:44:18 That's where the deep learning technologies are going to be able to provide us a huge value and that amplification identifying what users need to be watched, what behaviors need to be curved or alternatively, educate end users around risky behaviors that they could be faced with. All of those come together to solve this insider threat program, and all the technologies that we have are ultimately what's going to enable it.

Rob: 00:44:45 I will tell you this. I wish that there was a switch on the wall where I could flip it and turn off the darkness, but the reality is, right now, we're going to be continuing to face a new set of challenges every day, and only through great technology, great education, great culture can organizations make sure that they're kept out of the limelight for bad media that's coming out.

Sean: 00:45:11 Rob, I really appreciate your time on this today. Rob Juncker is Senior Vice President of Product at Code42. Thanks again for joining us.

Rob: 00:45:19 Thanks so much for having me. This has been great.

Sean: 00:45:24 To get a sense of how AI related technologies are being brought to play and to stop insider threats now, we spoke with Justin Fier, the Director for Cyber Intelligence and Analysis at Dark

Trace. Here's what he had to say about detecting insider threats.

Lee: 00:45:40 All right. Now, we are joined by Justin Fier, who is Darktrace's Director of Cybersecurity. Thanks for joining us, Justin.

Justin: 00:45:48 Thank you.

Lee: 00:45:49 I want to discuss since we've been talking about insider threats some of the particulars around AI and insider threats to tie all of this together. So, to start with here, what can AI and machine learning do, in general, to detect potential insider threat behavior? I know that's really broad, but we can maybe start there and drill down a little bit.

Justin: 00:46:11 Yeah, and I don't think it's as broad as a lot of the listeners might think. Machine learning, I think, is perfectly suited for insider threat just simply because anomaly detection, in my opinion, is the best route to detect it. An insider is always going to deviate from their standard pattern of life. They're always going to change what they do in order to accomplish their mission. Machine learning is just perfectly suited for that.

Lee: 00:46:41 That actually builds off of something that we heard from Kathleen Carley when we spoke to her earlier. She mentioned the same thing that when insiders begin to engage in potentially troubling behavior that they withdraw a little bit from normal associations or they change patterns. Generally, it's a withdrawal kind of behavior. Does that mirror what you guys typically see?

Justin: 00:47:03 Yeah. I think that's one of many. I think the other thing to think about is, oftentimes, insiders have guilty knowledge about how the internal infrastructure works. So, they sometimes know how to attempt to evade detection. So, even those attempts innovating detection are a deviation from their standard day-to-day work.

Justin: 00:47:27 So, I think in many of the cases that we've seen, it could be something as simple as just browsing a file share that you've never browsed before, all the way up to trying to send out larger amounts of data than you typically would in your typical job function.

Lee: 00:47:46 This sounds a teeny bit like the minority report style pre-crime reporting, which potentially could squeak some people out. So, I have to then build on this question by asking, how do you deal

with separating the proverbial wheat from the proverbial chaff here?

Justin: 00:48:03 Sure, sure, and that's funny. You hear the term false positive a lot in our industry. I think it's an antiquated term. It, typically, in my opinion, refers to the more legacy approach to security, which is rules and signatures, which I think we all understand does not apply to insider threat. There is no rule or signature for an insider threat.

Justin: 00:48:24 Really, in anomaly detection, a good model doesn't generate false positives. It generates positive positives. If it fires, it actually is an anomalous activity on the network. We see that all the time.

Justin: 00:48:38 Now, was it just somebody that deviated from their standard pattern because they had to work late one night or had to send out a large file or was it actually malicious? That's the difference, and that's up to the security team to ultimately decide after the model fires.

Lee: 00:48:57 So, I think you've answered the second main question I had here in my list, which is that with the growth of legitimate businesses using more and more cloud sharing, having more and more data flowing potentially off the land and out into the internet, and the ability to conceal a great deal of data exfiltration within encrypted web traffic, how do you distinguish legitimate behavior from furtive or malicious behavior?

Justin: 00:49:23 Yeah. So, first and foremost, everything I hope is encrypted in today's day and age. For Darktrace's sake, we don't really care if it's encrypted. We're a metadata shop. I can still show anomalies without breaking into the payload. I think the important thing that many corporations need to start considering is getting those corporate policies in place. Unfortunately, it still hasn't happened. It's the wild West.

Justin: 00:49:49 I go into a lot of clients and there is no policy on what's acceptable cloud usage and business units fire up SaaS applications and various different cloud storage applications without ever notifying security.

Justin: 00:50:03 A simple solution would be sign up for one, make it the appropriate policy, and then use machine learning and other tools to police that and verify what's outside of the approved use.

Justin:	00:50:16	I think the other thing a lot of folks don't realize is the laws governing this use. For instance, if an employee fires up their own private cloud storage and there is no policy about doing so, I think the question arises oftentimes of after that data moves into that private cloud storage, who owns the data? I think some would argue that it's the employee's now because the company didn't take those actions to actually set what's right and wrong for doing that.
Justin:	00:50:45	I think a good lawyer could certainly litigate that and get the data back, but in some cases, the minute that data goes to your private cloud storage, you are now the owner of it. I don't think many corporations consider that enough.
Lee:	00:50:59	So, you had mentioned policies, and I know that we have had in the past a pretty significant discussion around the role of technology and hard enforcement versus the role of policy and human enforcement. When you're looking at mitigating insider threats, is there, from your perspective, a preferred balance between taking a technological approach with system policies, GPS or whatever versus a rules-based human approach with organizational policies and soft rules, I guess? Is there a good mix between those? Because you can't do everything with either, right?
Justin:	00:51:35	Right. I think it's a 50/50 mix. I think you really have to address both of them. I think the biggest shortcoming that I see in a lot of corporations, and it really is the lowest hanging fruit, is just integrate more with your human resources department. We've all taken certified ethical hacker, and we know the number one answer to what is an insider threat is disgruntled employee.
Justin:	00:51:59	Well, HR should be able to tell you that. They're supposed to know the employees better than anybody. I just don't see enough HR departments integrating with the security team in order to identify these people either before they become a risk or after they've already become a risk.
Lee:	00:52:18	I'd like to point this out also and have you address it when we're discussing insider threats, that's a very malicious and pointy sounding thing, but it's not necessarily always a purposeful kind of thing. People don't always set out to be an insider threat.
Lee:	00:52:35	I guess the canonical example is the, and I ran into this, which will show you the age that I am and when I was last working in IT, when modern smartphones began to get really, really big shortly after the iPhone came out and shortly after Android made its first appearance. We ran into a lot of instances at

companies I was at where the executives would show up with these brand new smartphones or the CIO would show up or the CEO would show up and be like, "Make my iPhone work with your email now," which leads to a tremendous amount of integration issues and unanswered security questions, right? So, how do you address insider threats that come potentially not from the bottom up from line employees, but from the top down as it were?

Justin: 00:53:20 One thing that we've started doing with a lot of our clients that I really enjoy and I find fascinating is building insider threat hunt teams. We've all heard of the cyber hunt teams, but an insider threat hunt team is a little bit different. As you mentioned, it's the technology and the human psychology element as well.

Justin: 00:53:40 One thing that we've implemented, instead of only looking at the high risk users, we also look at the high value users, so your entire C-suite, your executives. So, for instance, within Darktrace, a lot of those insider threat hunt teams will go in and tag those high value targets, and they'll watch them a little bit closer not because necessarily they're being a nuisance and accessing things on devices that are not approved, but also because they are higher value targets, and they typically tend to get spearfished more, and they tend to get targeted more as when they're traveling, et cetera.

Justin: 00:54:17 So, I think just taking the opposite approach of looking for the disgruntled employee also has worked quite a bit in our customers. Focus on those high value targets and just look at them with a little bit of a thicker magnifying glass.

Lee: 00:54:33 So, building on that, how can AI play a role in helping both, well, I guess on the hunt team side and also with helping employees learn compliance with policies that prevent accidental insider exposure?

Justin: 00:54:46 Oh, yeah. Absolutely. I mean, you've got the whole, and it's been written about for years, the unintentional versus the intentional insider. Again, I'll sound like a broken record. It comes back to anomaly detection. So, if I've baselined the entire network and I have a sense of what self is, I will be able to identify a finance business unit, for instance, firing up a SaaS application, sending out gigabytes and gigabytes worth of data.

Justin: 00:55:12 Now, it might've not have been intentional insider threat, but unintentionally, they're adding risk to the business by not vetting it with security, by not verifying that those channels that

		the data is leaving are secure, making sure regulations are being followed and they're not breaking any regulatory laws.
Justin:	00:55:31	So, it all comes back to that visibility paired with that anomaly detection. I think a good baseline, a good set of models will make even the tiniest deviation light up like a Christmas tree.
Lee:	00:55:45	You mentioned my favorite word, risk, because risk has a very special meaning if you're a CIO or a CTO, and an InfoSec. Risk, it has a very specific industry meaning. When you are coming in to discuss with companies about Darktrace and going in and setting up your baseline or whatever, do you have a lot of discussions in terms of risks specifically?
Justin:	00:56:05	Not necessarily in the first discussions. It's usually after we got the box deployed and during our proof of value. A company gets to see all these many blind spots, whether it's all of the IoT they didn't know it was on the network or all of the different cloud applications they didn't know was on network. That's usually the big eyeopener and that's when the discussions start being had, "Well, man, we need to start having some policy discussions about these things."
Justin:	00:56:32	As I said before, I think visibility paired with the anomaly detection is hard to beat. Unfortunately, even in 2020, we still lack a lot of that visibility into our network. It's sad to say, but many of our customers are still blind to almost 25% of their network, and that's all the nonstandard, traditional IoT devices that are out there, and we haven't even begun to think about what 5G is going to do to that.
Lee:	00:57:01	So, I was wondering if you could maybe walk me through what, hypothetically, what a company might see when they first began to do an assessment like this and set up a baseline. So, can you give me a couple of examples of the types of behaviors, and threats, and risks that insiders might exhibit that companies might not be aware of that you guys spot when you come in and spin up your solution?
Justin:	00:57:25	Yeah. There's dozens. So, I'll pick a couple of the more interesting ones. So, it typically takes us anywhere from seven to nine days to baseline. After that seventh day, we've got enough data that the math can start to kick in and pop up some of the unusual activity.
Justin:	00:57:41	Now, we've seen a lot of cases of raspberry pies being inserted into the network for malicious reasons. Now, I have nothing bad

to say about the raspberry pie organization. I've got dozens of them around my own home, and they're a client of ours, but we are seeing them in the ceiling tiles, under the floorboards really doing creative ... Here's a nerd reference for you, Mr. Robot type attacks within the corporate environment.

Justin: 00:58:12 So, I'm blown away by some of the things I continue to see just with those devices. Part of that is because it's a $5 device that I can buy locally. We're also seeing a lot as we've already talked about in this talk of just cloud exfiltration, but I've seen some more interesting cases where people have tried to exfil data out through the Apple instant message program by creating their own account on the other end and sending file attachments.

Justin: 00:58:45 To most security teams, it doesn't look anything more than encrypted, iChat communications, but when you start looking at it from a metadata perspective, you start to see an increase in packet size. You start to see a steady pattern of life. So, we were able to detect pretty quickly that someone was actually sending out files at a pretty steady pace.

Justin: 00:59:08 Then there's the just standard practice person goes into HR, gives two weeks notice, walks right back to their desk and fires up an EC2 node or a digital ocean node and just starts exfiltrating data out of the network. Again, despite the fact that it's 2020, many corporations still lack the ability to spot gigabytes and gigabytes of data leaving the network. Just because our networks are so complex today that it's hard to differentiate that little needle in the haystack.

Lee: 00:59:42 Building on that, and this is a broad question, but do the insider threaty style trends you guys see at the micro level when you go into companies, are you seeing this echoed perhaps at the macro level across industries and across countries?

Justin: 01:00:00 I don't think we've seen that yet. Unfortunately, we're writing up dozens and dozens of insider threat cases per day. I have not yet seen where one industry is more dominant over another industry. The unfortunate cases, if you just look back at the last few years of the publicly documented insider threat cases, it covers everything from the media industry to leaking up and coming TV shows, to the financial industry if you look at the Capital One breach. I think every industry is victim and will fall victim to this. I think it's one of those target of opportunity events.

Lee:	01:00:45	All right. Well, thank you for taking the time, Justin. This has been Justin Fier, Director of Cyber Intelligence for Darktrace. Thanks a lot, Justin.
Justin:	01:00:52	Thank you.
Sean:	01:00:54	So, AI may help in identifying signs of someone acting irregularly, but whether or not it can reach the level of a precog like in minority report, recognizing pre-crime is open to debate. For now, it still requires humans to get that insight into another human's intents. Next time, we'll talk about another area of artificial intelligence research that is growing ever closer to reality, adversarial AI. I hope you'll join us.
Lee:	01:01:23	Once again, this episode was sponsored by Darktrace, the world's leader in AI cyber defense. With more than 3,000 organizations relying on its AI technology around the globe, Darktrace is transforming security from the inside out. Start your 30-day free trial by visiting darktrace.com/trial.