**RedZone Podcast Episode #47: How to Make Better IT Security Investment Decisions – with Marshall Kuypers**

Bill: I'm very interested in risk and in educating and spending time talking with senior decision makers and helping them understand some of the thought leaders, in risk analysis. I thought that we could spend some time going through your research and some of your findings in particular, related to practical quantitative risk analysis in cyber systems. How did you get interested in this topic overall? What was the genesis of it?

Marshall:
[00:01:00] My background is actually not really in cyber security. I'm fairly new to the field and I got my start in risk analysis more generally. I'm currently working on my PhD at Stanford University and the other folks in my research group study, quantitative risk in a number of other fields. For example we have somebody who's studying the risk of satellite failures, we have somebody looking at the risk of the failure of the power grid, failure in medical devices.
The application that I got interested in happened to be cyber security. One of the things that we noticed when we talked to a lot of decision makers was that, while many other types of business risks are fairly well understood. For example workman's compensation or how often a building catches on fire or something like that. Cyber risk is really not very well understood and so often you'll have decision makers, trying to quantify cyber risk using very qualitative scales.

[00:02:00] They'll talk about things being green, yellow, or red. These qualitative assessments are informing really big decisions of millions of dollars at a lot of organizations. What we wanted to do is try to bring an increased level of rigor to that so that we could help organizations, ask questions about how effective those different security safeguard investments were and where they should be putting their money.
Bill: I think that that's a great start. Would you say that there's significant uncertainty around cyber security investments right now?
Marshall: Absolutely. From a high level it's really difficult to know if you're an organization, where you should be putting your money. If I tell you I'm going to give you a million dollars to improve your security budget, that might go into better encryption technologies, or a better firewall, or maybe even employee training, to teach your employees how to not click on phishing links. Right now it's really difficult to know where that money should go or the corresponding risk reduction associated, with any of those investments.
[00:03:00] This problem is really exacerbated by the fact that there's just a ton of security vendors that are out there, that will sell you pretty much anything under the sun. There's a lot of snake oil and a lot of security products that just don't really deliver. Again, what we're trying to do is figure out a way to quantify those cyber security investments more effectively, so that we can help organizations secure their environment.
Bill: Let's talk about this. When you were doing your research, did you actually physically have to go interview companies and like how many companies would you have to have examined, or did you just look at raw data that was public? How did this morph and evolve as a project from the beginning?
Marshall:
[00:04:00] That's a great question. This turns out to be a really difficult thing to actually get data that you can do high quality research on, in cyber systems. In my opinion this is one of the reasons that we haven't seen more progress in cyber security research over the last 15 years. Cyber security data is just tremendously difficult to get. When I say cyber security data, I'm specifically referring to data on cyber security incidents.
I'm really interested in talking to an organization and knowing how many times laptops, have gotten stolen in the last year or how many websites have been defaced, or how many malware infections have you had on user's endpoints. Most organizations actually track this information in internal auditing systems, maybe even in Excel spreadsheet but, most organizations usually have a fair degree of visibility into that.
I'm not really interested in log data, I'm interested typically in these cyber security incidents where an investigator will go and open up one of these incidents, they'll resolve and investigate it and then they'll close these incidents. Even though organizations have these data, it's really difficult for researchers to get.
[00:05:00] Basically we tried to use every trick in the book where we would talk to different organizations, and develop a good relationship with them where we could say look we'd really like to take a look at these data, because we think that there are really meaningful insights in them that we could discover. What we found was that when we talked to the vast majority of organizations so even huge tech firms in the Bay Area.

A lot of these organizations reported these data of how often they see different types of cyber attacks, but there were very few companies who were actually doing any analysis on these data. What we could do is go in and we could say, “Look if you give us access to these data, in a very short period of time like we're talking about a day in Excel, you can find some really meaningful insights to dramatically improve your situational awareness.”
Once we've been able to demonstrate this with a couple companies there have been a steady stream of other organizations, that have expressed interest in collaborating with us.

Bill:
[00:06:00] You're speaking at some significant events. I found it interesting that a young researcher like yourself is actually gone right to the top of the heap, so to speak from the visibility point of view. You must be onto something important because what's interesting is that as I have these conversations on risk with Jack Jones and others, the probability and likelihood data of events happening seems to be the missing element. It's one of the more challenging ones to gather information on.
You're saying that you can actually gather this organizational data but you're actually going direct to companies that are, storing this information for mainly auditing purposes but they're not necessarily making the information public?

Marshall:
[00:07:00] Exactly. There are very few incentives for organizations to disclose a lot of these data. For example, like media attention could very easily drive one of these company's names through the mud and unfairly make them look like they're doing a poor job at cyber security. Some of these are even definitional problems. They're probably well known to many of your listeners where if you say that an organization has experienced, a million cyber attacks in the last month the public may see that as extremely alarming.
Whereas any security professional will say, "Well, if you're talking about millions of attacks, then you're probably defining a port scan as an attack." Which is not very interesting. It is very difficult to get these data. Historically researchers have had to rely on publicly reported databases. There are a couple pretty high quality ones that are out there. For example the Privacy Rights Clearinghouse is a great resource.
[00:08:00] The VERIS Community has a great database that people can go and look at a bunch of cyber security breaches, that have been recorded over a 6 to 10 year period of time. Those are still a very biased view of the certain incidents that an organization might be experiencing. There are really only 1 or 2 if any research papers that look at all cyber security incidents, at a single organization.

Inherently if you rely on those publicly available databases you're going to be only looking at those really large impact incidents, and you're not really getting a view of those smaller impact incidents. What we've found when we actually go into an organization, we see everything from start to finish. Low level attacks but then also those very severe attacks. We find that the organization actually has some patterns in how it experiences attacks that, are tremendously useful from a risk modeling perspective.
[00:09:00] Number 1 it turns out that a lot of the rates or the frequency of these cyber security incidents, is remarkably consistent over time. A lot of us will read reports like from the Verizon Data Breach Investigation Report or other industry reports, that basically indicate that the number of attacks is going up dramatically and the number of cyber breaches is also skyrocketing. There's actually virtually no evidence that that's actually true.

Even in the publicly available databases, there have been several academic papers that have been published that show that the rate of attacks, is remarkably consistent over time. Then similarly they've looked at the severity distribution. How bad are these attacks? Are they getting worse? Again I think if you talk to most laypeople they say, “Absolutely.” Cyber is increasing in severity, things are getting much worse. We're headed towards a really dark period.”

[00:10:00] Again there's virtually no statistical evidence to support that. All of the numbers that we actually have when we actually look at them, show that the severity distribution of a lot of these cyber attacks is also remarkably consistent over time. What we've basically shown is that by again looking at all incidents at a single organization, you can start to draw out these really interesting and sometimes counter intuitive insights, that can really aid a decision maker and how they're doing resource planning.

Bill: This is really fascinating. We'll start with counter intuitive insights. We have one way of gathering the data which is looking at the Verizon Breach Report, you can look at some of the Clearinghouse information, and such. That's common knowledge. If someone's willing to do the research, they can dive in there and make some assumptions, and probability, and events, and things of that nature happening over time to fill in their probability and likelihood gaps.

Let’s take your approach though and your method and dive into, because I really want to get into this counter intuitive findings and patterns that you're finding.

Marshall: Can you repeat the question there one more time?

[00:11:00]
Bill:
Sure. If we have to get into some of the counter intuitive things that you're finding. We know the data exists and the models exist and the risk is quantifiable. We know that there's some of the assumptions you're finding are false or misleading or don't exist, but you're finding some different evidence. I'd love to hear more about that that would seem counter intuitive to a normal way of doing it.

Marshall: Sure. I think that absolutely that the first major insight that we've been able to determine is, the fact again that the rate and the severity of many of these cyber security incidents is not changing nearly as rapidly as most people think. Again, we find some weak evidence for example that malware attacks are going down at certain organizations, and that some attackers may be pivoting to using certain attack factors like ransomware or malicious email more frequently.

[00:12:00] What's really interesting is that for a larger organization if you're to look at the data, what you would probably find is that this change is occurring incredibly slowly, like on the timescale of years. The cyber domain is evolving but again not nearly as quickly as a lot of other folks think. This is a really important consideration.

I think one of the reasons that the field of cyber security professionals, may not be as interested in using historical data is because, they think that everything is changing so quickly. That it's really difficult to get any insights from something that happened 2 or 3 years ago. While cyber security is undoubtedly a attacker-defender problem where there are, dynamics and evolution that are occurring. Again it's very interesting and surprising that this is actually occurring on a fairly slow timescale.

What happened last year at your organization is actually a fairly good indication of what might happen in the next year.

[00:13:00]
Bill:
Malware for example is what you're saying is that the media is talking about malware and its different variants, from Shellshock to CryptoLocker to all these different exotic forms of malware. What you're saying is that although we might classify these as exotic and increasing, the actual rate of these incidents is remarkably steady.

Marshall: It would be really interesting to hear from many of your listeners if this fits their intuition. While there has been a lot of media coverage of some of these more interesting cyber attacks, in the last couple of years my intuition is that these have been happening for the last 10 or 15 years. That most security professionals would say, “Yes, we've been dealing with attackers at a steady rate for a long period of time. It's just that the media wasn't that interested in them until a couple years ago.”

[00:14:00]
Bill:
Yeah I would agree with that. I think the interesting piece here is that when you're developing the likelihood and probability of events happening, it's very easy to be swayed by public opinion versus evidence. I think the amount of time and effort that goes into gathering evidence is potentially daunting for folks. If you're a layperson, CIO, or CISO that wants to get to the bottom of this quickly, how would you recommend that they do this?

Marshall: This is a great question and it's something that I would love to talk about. To take a step back my research is really focused on normative decision making or in other words, how should people make decisions about investing in cyber security? Before we figure out how people should be making decisions one of the first steps is to figure out how people currently are making decisions, or the behavioral side of things?

[00:15:00] There's a really rich interesting field of behavioral decision making that I think is almost one of the, fundamental pieces of background knowledge that security professionals should have. For example there's a ton of really great research that shows that people are really bad at making decisions, and assessing probabilities and doing all other things. For example, there's evidence that shows that the order in which you ask a question can change the answer that somebody might give you, or the way that you phrase a question.
[00:16:00] For example people have gone out to doctors and they say, “How many of you would recommend surgery in the case where the 1 month survival rate is 90%?” Then in the 2nd case they go to another group of doctors and they say, “How many people would recommend surgery if there's a 10% mortality in the 1st month?” You can see that 90% survival rate and 10% mortality that's the same thing, it's just phrased differently.

In the 1st case where you're highlighting the fact that 90% of people survive, 84% of doctors recommend surgery while in the 2nd case when you highlight the mortality, only 50% of doctors do.

Bill: Yeah that's funny.

Marshall: It’s terrifying to a certain degree because if I'm going in for a major surgery I don’t want the way that I phrased the question, to impact the recommendation that my doctor is going to give me. Again, there is just a huge body of scientific evidence at this point that shows that the way that the way we phrase questions, absolutely changes on how we think about some of these answers. There are a bunch of other interesting names for these types of biases as well.

[00:17:00] For example, there is the Recency Bias. For example if I am biking around campus and I have somebody who comes screaming, and there’s are an undergraduate and they almost run into me. They are not wearing a helmet, they almost cause me to crash. That’s really fresh in my mind. For the next week after that I’m going to be a really defensive bike rider, because that event happened very recently. However, as time goes on maybe a month later I've forgotten about that.

Maybe then I start riding my bike much faster than I should. This is something again that we really see in cyber security. After the Snowden leaks there were a ton of chief information security officers, that were really interested in trying to find their malicious insiders. Even though in many organizations there's virtually no evidence that malicious insiders are actually a major threat to that organization.
[00:18:00] We have to really aware of these biases that we have inherently, in order for us to make better decisions moving forward. If your listeners have some free time I think one of the best resources for this is a fantastic book called, thinking fast and slow by Daniel Kahneman. He is a Nobel Prize winning economist. It’s a fantastic, really interesting book. It’s very accessible. It goes through basically all these biases that human have.

Once you start to learn about those and understand them it really arms you so that you maybe can’t necessarily correct for them, but you know when you might be making a biased decision.

Bill: I think this is a stunning point you’re making important. By the way I’ve read that book as well it's really wonderful for people to get. I agree with you. It’s so much based on fear, and uncertainty, and doubt, the profession. It's really easy to just get the fear part of the brain really going and really forget some of the logical empirical data and evidence, because and we're completely acting from fear which is clouding intuition and our decision making. You also made the point of parole decision.

[00:19:00] Maybe you can explain the parole decision making for Israeli prisons?

Marshall: Sure. There is a really interesting paper that basically looks at the proportionate favourable decisions, for people who go in front a parole board if they're in an Israeli prison. It shows basically what there likelihood is of getting parole throughout the day. What you notice is that it starts pretty high and then there is a steady decrease throughout the day, of the chances of you making parole. Then there are 2 huge jumps and we can look at that.
[00:20:00] We can be like, “Well that’s really strange. What could be the causing that?” What the authors of that paper found was that those 2 jumps were all of a sudden, basically like if you’re 15th line you have only a 10% chance of making parole. If you’re 16th in line then you have a 80% chance of making parole. Those jumps actually corresponded to lunch breaks. Really terrifying that the judges who are making incredibly important decisions who are supposed to be impartial, and they have great judgment are being influenced by their blood sugar levels.

If you happen to find yourself in an Israeli prison make sure that you go in front of the board right after, everybody has taken their lunch break. That s you’re big take away from today. Your point earlier that we have all these biases and we really need data to inform these decisions, I think that’s absolutely correct. With 1 chief information security officer that we worked with we asked them, “Where do you think your major risk is for next year? Where should your security budget be going?”
[00:21:00] They said, “Absolutely malicious insiders. That’s what I’m really worried about, that’s where I think all my risk is.” We looked historically at this organization. They are a very large organization. Over a 6 year time period they had exactly 1 malicious insider incident. This was somebody that they could have seen coming a mile away, they were a system administrator that had gotten fired. The organization didn’t revoke the login credentials. Later that night once they had been fired they’d logged back in and messed a bunch of stuff up.

During that same 6 year time period they’d over 200 website compromises. These ranged from website defacements, SQL injections, they got huge amounts of data. Then also included nation states that were actively compromising and exfiltrating large amounts of intellectual property, from this organization. When we show this to the chief information security officer it was really an aha moment where they said, “Wow.
[00:22:00] My intuition really does not correspond to the reality that the data is showing us, of where the actual risk is in our organization.” Again once we know that this really demonstrates the importance of needing to go out and base our decisions, in actual hard evidence and data. If we use these qualitative methods or hand waving methods there are just so many ways that our intuition can fail us, that we end up making really bad security investment decisions.

Bill: Did you use the word hand waving?

Marshall: Correct.

Bill: Basically meaning that someone’s waving their hands because this is such an important event and they’re waving their hands.

Marshall: I would almost say waving your hands and you’re making it up. You’re talking in a very overt way so you’re saying, “Oh trust me I’ve thought about this and this is really what we need to do.” Lots of times you’ll see security vendors doing this where they’ll say, “Buy my product it will solve all your problems. I really understand what your core business need.” We want to be able to test that.
[00:23:00] We want to be able to say how many of the attacks that we see actually coming in through our email and what’s a low medium high estimate, of how effective this new email filter might be. Then let’s take a look at the different impact measures for these different cyber security incidents, and let’s see which are most costly. Let’s base our decisions off of those instead of just what somebody is telling us.

Bill: Let’s get super practical. One of the things you had talked about was the data exists. You were saying you weren’t necessarily looking at log data which is interesting. You were looking at obviously other pieces of data. What are those types of information that if someone’s listening they can merely go talk to their teams and start siphoning the data points?

Marshall: Again I’ll make the distinction between log data and security incident data. Log data I think can be useful in some context especially if you’re trying to piece together exactly, what happened in a certain cyber security incident. The issue with them is that there is a tremendous amount of noise in those signals. You’re dealing with a huge amount of data and it’s really often difficult to know what’s going on. Further to really find any actionable information from those log data.

[00:24:00] Really at the end of the day what we’re most interested in finding are data that we can make actionable decisions off of. What I encourage organizations to do number 1 is if they’re not recording cyber security incidents, they should start. When I say cyber security incidents again it goes high level incidents. How many times a laptop gets stolen. Again every time you see a website attack or an SQL injection where somebody actually makes off with information, malware infections, maybe somebody is sending a phishing email into your organization.
[00:25:00] Any of those real incidents where somebody from the security operation center or the security operations team, has to go in and investigate. Lots of times these will involve the attacker actually getting some information. Sometimes it may be just ensuring that the attacker was unsuccessful. For example if you have spam emails coming into your organization then your investigator is going to go in, and make sure that nobody clicked on something that they shouldn’t or nobody exposed their credentials.

If they did they’re going to have to reset those credentials. I would say focus on those high level incidents. Then start recording those if you’re not already. I think that the things that are most important to record are number 1 these incident categorizations. You can do this in the form of tags. You can say #emailandsender, #websiteattack, or something like that. Then the other major thing that’s really valuable to record about these incidents would be, some quantitative impact measure.
[00:26:00] Historically this has been difficult to do. Lots of times if there’s even personally identifiable information that’s disclosed, maybe you might be able to record the number of records that are exposed. Even that has some uncertainty. Then what do we search again into other loss categories like reputation damage. That’s been so difficult to quantify historically. A lot of organizations just say, “Well I don’t know if it’s between 10 million or 30 million.

I’m just going to call it high.” There’s actually a huge amount of value in specifying that it's between 10 and 30 million. We find that a lot of these cyber security incidents follow what’s called a heavy tailed distribution. Which basically means that you have incidents that span the severity across orders of magnitude. The vast majority of incidents at your organization are going to have a severity or cost of less than $100.
[00:27:00] Then every so often you’re going to have 1 that costs your organization $10,000 or $100,000 even millions of dollars. When you have that spread across orders of magnitude that’s really the hallmark of what these heavy tailed distributions are. Knowing that you’re in a heavy tailed distribution is tremendously valuable for a decision maker. We can talk about that in a little bit if you’re interested. Basically you'll need some quantitative metric to record to be able to determine that.

Again if you’re an organization and you’re recording these cyber security incidents encourage your security team to put down perhaps how many hours of investigation they spent, investigating and remediating an incident. We’re not really interested in the difference between 7 or 8 hours or 32 or 33 hours, what we’re really interested in is the difference between a 5 hour incident and a 500 hour incident.

Generally folks are pretty comfortable getting the order of magnitude right. If you get your order of magnitude right and you have some quantitative measure instead of just saying this was a high, medium, or low or a tier 1, tier 2, tier 3, incident. There’s so much more that you can do with those data.
[00:28:00]
Bill:
The value of that is just so that you can start to be able to show decision makers the potential impact and what it cost the company. Is that one of the pieces you’re trying to get to?

Marshall: Absolutely. Again at the end of the day what we’re trying to do is actually come up with a final number for cyber risks. We want to tell the organization, “This is how much we think you’re going to lose next year.” Instead of just coming up with a point estimate we’re actually going to give them a probability distribution. We’re going to say, “There’s a 30% chance that you’re going to lose less than a million dollars due to a hack, and there’s a small maybe a 1% chance that the next year is going to be really bad.
[00:29:00] Then you’re going to lose 10 or $20 million.” Giving that whole probability distribution is where we can really start to inform a better decision making process. The CSO can take that information to the board and they can say, “Look here’s our risk curve that we have currently. Then if we put a new firewall in place or a new intrusion detection system this is how that actually changes that.” Again we’re trying to get towards that quantification.

Being able to record the severity of different cyber security incidents basically allows us to come up with those inputs, that we can put into this risk model. There are a bunch of really interesting mathematical subtleties for why this is important to do. Going back again to the fact that a lot of these incidents follow a power law distribution. What that often means is that even though you’ve only had incidents that range between 1 and 1000 hours of impact, you can with high confidence extrapolate into how often incidents of larger severities will occur.
[00:30:00] Even though you’ve never had an incident that takes 10,000 hours to investigate or remediate the fact that this is a power law distribution or a heavy tailed distribution, often allows a decision maker to be able to say, “Based on the information that we have we think that we’re going to have one of those major hacks that cost, 10,000 hours of incident investigation time. We think that’s going to occur once every 10 years or something.”

Bill: I can see why this really helps. I was looking at some of your examples during our conversation about lost devices. We can break this down into it supports the security investment in a major way. If you’re going after a $10,000 investment or $100,000 or a $1 million investment in some IT security object or service, and you can then back that out to the likelihood of an event happening. You can map it to likelihood and risk.

This is raising the level of sophistication more to what finance is used to, what insurance industry is used to. It’s raising the bar from the IT security perspective.

Marshall:
[00:31:00] Absolutely. I'll have to say that the cyber insurance folks are some of the most interested people in these new tools. They understand lots of other types of risks very effectively. They have a general intuition for how much they should charge me as a young male for a car insurance, based on demographics and things like that. However for cyber security we basically have none of that information.

Cyber insurers really have no idea if the size of the organization matters or your website security actually makes a difference, or what domain you’re in matters either. What we want to do is get a general intuition for this to be able to test some of those assumptions, and to be able to get and quantify risk instead of relying on low, medium, high, severity assessment. I think that one of the main values in risk quantification or of the CSO would be, how do you actually communicate that risk to the board level?

[00:32:00] Currently if you going into a boardroom or you’re talking with your chief executive officer and you say, “Look I want $1 million to pull this inscription on all of our devices.” There are a lot of boards that will say, “How much money is that going to save? What is the cost effectiveness of that investment?” Right now there aren’t a lot of tools out there that can help a CSO answer that question.

If we actually do risk quantification techniques we can say, “Oh well it turns out that it’s going to cost just $1 million to pull this inscription on all of our devices. We think that we’re going to save on average $2 million per year based off of these assumptions.” This is a really great investment. It’s a 4 to 1 benefit cost ratio. This helps justification for a CSO’s IT security budget which we think is a really good thing.

Bill:
[00:33:00] Being able to take this down to that summary level with all the support data underlying it takes the emotion out of the decision. It takes the raw intuition and the raw emotion. I think intuition is always going to be there. You can’t make it go away but we can certainly take the emotion out of it. We’re just dealing with more I would say adulterated regular business decision making.

Marshall: I agree. If you look at most other boardrooms almost all other business risks are assessed on a monetary scale. It’s interesting that even in some of the financial services industries they would have these risk registries, for how risky or what the probable losses are for several different incidents that might impact the organization. Historically they used dollars for everything except for cyber which is really bizarre.
[00:34:00] They’d have this other scale maybe they talk about the confidentiality, availability, integrity, of data but they don’t talk about dollars. What we really care about is how much are these cyber security incidents going to cost us? While this was a really difficult question to answer 10 or 15 years ago there’s actually a surprising amount of high quality research, in the industry reports and data these days that can inform those questions.

We can actually help organizations put a monetary value on a lot of these things, like reputation damage or the loss, of intellectual property, or other like equipment damage, that historically have been very difficult to quantify.

Bill:
[00:35:00] That includes PII fines, investigation costs, forensics cost. All that are you saying is it in a single source or is it something that’s just a couple websites, that people can go to? Where would people go to to find that information? You call it data spillage. Is that what you use to refer to the categorization of when you lose data and information what’s the cost? Is that something that’s a subscription service or something that’s online publicly available?

Marshall: I would say that the information sources for these are all over the place. What I’m trying to do for my dissertation is basically provide a cookbook for how an organization can go through and compile that information, and basically present a comprehensive review of what all those data sources are and why they’re valuable. Some of them you do have to look pretty hard for. For example legal fines due to cyber security incidents is a obscured topic.

[00:36:00] There have been a couple academic papers that have been written about it. There you really have to know where to look to actually get good probability and impact information. The chances that somebody is going to sue you if you have a data breach. There are other source of information for other types of data losses that are really accessible. For example investigation time. If an organization is recording this already they have that data in-house.

Similarly for the rate of different cyber security incidents. How often a website defacement occurs. This is something that a lot of organizations have in-house. There are other vectors too, other attack impacts. For example reputation damage. Here what we are basically trying to help organizations do is quantify reputation damage, via an elicitation. It’s called a willingness to pay or expert probability elicitation. Here reputation damage is highly dependent on what type of organization you are.

[00:37:00] If you are a retailer or you’re a defense contractor or you’re a public university. What we'll do is we survey the literature and get an idea of how many of these incidents have occurred before. For example we might be able to look at the target data breach or the RSA hack. There those companies are still in business. Intuitively we know that these cyber security incidents or the reputation damage, they’re not costing that costing that organization billions of dollars.

[00:38:00] Then we can also look at some of the numbers that are reported in the media and it’s certainly going to cost that organization more than 50 million, if you disclose on the order of a couple 100 million on credit cards, personally identifiable information records. What we’ll do is we’ll actually work with the organization to try to quantify that for the organization. The way that this works is we sit down with the decision maker and we'd say, “What is the reputation damage that you might experience if you have a really impactful hack?”

Often times a decision maker will say, “Well I have no idea.” We’ll say, “Really? Do you think it’s more than $10?” They say, “Well absolutely of course it’s more than $10.” Then we'd say, “Well is it going to cost your organization less than 10 billion?” It's like, “Well of course it is.” We say, “Great. Well we’re on our way to quantifying what that reputation damage is.” What we basically do is we try to talk with them to get a range of reputation damages.

Basically what we do is incorporate that entire distribution or that uncertainty over those different possible losses into the analysis. Again we’re not looking for just a point estimate. We don’t want a decision maker to say, “Well if we have a really bad hack it’s going to cost us X.” It’s totally fine to give us a range. While we incorporate that range into the models that we’re creating it turns out that that’s, a really important feature of being able to make well informed security decisions.

[00:39:00] You don’t want to assume away that uncertainty. You actually want to embrace that uncertainty. You want to incorporate it into your models.

Bill: I like your concept of embracing uncertainty because I think that is counter intuitive for the security profession. I think it makes a tonne of sense the way you talk about these large events or not outliers. I really like that concept.

Marshall: Absolutely.

Bill: As far as someone that can get started just to summarize basically we can start to look at our internal data, the incident data, the time to resolve the information, and just start to tag the types of incidents as we see them. That’s a very practical starting point to get started right?

Marshall:

[00:40:00] Absolutely. My number 1 recommendation again for organizations is to take a look at the data that you have, and then try to run some analysis on it. You don’t need a masters student from a top university to do this. This is something that an undergraduate in statistics should be able to do. They should be able to do it in Excel. Many of these organizations that we’ve worked with it was surprising to us that this data existed, and nobody had ever looked at them.

We could go into an organization and in literally 3 or 4 clicks in Excel, we could put together a graph that would just blow their minds. We could say, “Look it actually looks like the rate of cyber security incidents at your organization is pretty steady.” Or at times, “Except for this little blip right here back in 2013.” They say, “Oh yeah. That of course corresponded to an attack campaign that we experienced.”
[00:41:00] Absolutely encourage your folks to go out and just do some base level statistical analysis on, these cyber security incidents. If you’re not currently recording those that is I think one of best things your organizations can do, to improve it’s situational awareness. Start coming up with some categorization or you can consult some resources like the various community, to look at what their cyber security incident categorization is. Just start recording data.

Then a year down the road you’re going to have this really powerful tool that you can go back to basically look to see, what is occurring at your organization.

Bill: What I love about the way you bring this together as we come to a closing point here Marshall is that, the VP of sales when he walks into the CEO and says that this is the sales that we’re going to deliver to the organization this year, it’s never these are the exact sales there’s a range. There’s some uncertainty and there’s a lot of discussion about how that uncertainty is going to be addressed, mitigated. There is a high watermark, low watermark. It’s got to be about the ranges.

[00:42:00] There’s certainly bare minimums of the sales that the company needs to deliver on. We’ve got to get out of IT security of certainty. There is going to be a range and if we start to communicate it to the board then we can make it align our investments, to address uncertainty from this point of view.

Marshall: Absolutely. I’ll add there that it’s not just that we need to give ranges. There’s distinctions here that giving a color like green, yellow, or red, is not the same as giving a range. A range inherently has some uncertainty with it. It’s a quantitative expression of uncertainty. You’d never see a sales person go in and say, “Well we think that sales next quarter are going to be yellow or orange.”
[00:43:00] That’s meaningless. They actually put some uncertainty in that. They put a range in but that range is quantified to a certain degree. That’s exactly what we need to do for information security as well.

Bill: Thanks for making that point. Other than the book that you recommended earlier that you had said Think Fast and Slow. What other books have you come across that you find either really useful from the behavioral point of view that you were making earlier, or [inaudible 00:43:20] or from the just raw statistics that you’ve run into?

Marshall: I said that there are a tonne out there in my opinion. I think security professionals actually need to be really careful about what they’re reading and what they’re consuming. For example there are a lot of industry reports that are out there that in my opinion are worse than useless or even harmful. Where you have industries that are using totally bogus definitions of words like malware or virus. In some cases you even have industry reports that will go so far as to make up statistical words.

[00:44:00] They’ll say things like precision intervals which don’t actually exist. There are confidence intervals, they're credible intervals, but precision intervals do not in fact exist. You have to be really careful about this in practicing. I don’t necessarily have a good list of resources for people to go out to consult. I think that the book Thinking Fast and Slow is absolutely a really great resource. There’s a similar book written by Malcolm Gladwell called Blink that some readers may be familiar with.

It’s of the same flavor as Thinking Fast and Slow but I’m actually going to strongly encourage folks not to read Blink and to read Thinking Fast and Slow instead, for reasons that I won’t go into. I think that there are a couple of other resources out there as well. Some people may be familiar with the company RiskLens or the fair method or the Fair Institute. This is some work that’s been put together by some folks that is along the same lines of all the things that I’ve been advocating for.

[00:45:00] Risk quantification, uncertainty, data driven analysis. It’s some really high quality stuff. If you are interested in a good book for this type of analysis your readers can go out and find this book online. Give me 1 second and I’ll-

Bill: Actually you talking about the Jack Jones book the Factor Analysis and Information Risk book?

Marshall: Absolutely yeah. I would really recommend Measuring and Managing Information Risk: A Fair Approach by Jack Jones. I think in my opinion is a really great resource for security professionals.

Bill:
[00:46:00] I’m glad you brought that up. He’s done the podcast a couple of times. I agree with you. That’s why you’re here too. I feel this education is what the profession needs right now to bring the standard that we deliver to the board, to bring the high watermark up even higher. This is really good. One of the things I want to talk about just as we go your interest in the complexity group at Stanford. I found that incident interesting. I know there’s not necessarily a direct correlation between what we’re talking about in it.

I'd love for you to just give an overview what the complexity group is at Stanford and some of your personal interest in it. Just if you can give an example for everyone listening of what it is. I think that will be a good way to wrap up.

Marshall:
[00:47:00] I’m the co-president of the Stanford Complexity Group. We study complex systems. Complex systems again this is going to be fairly distinct from my research in cyber systems. I think it’s tremendously interesting and actually a really important concept for people to get, which is why I’m involved in this group at Stanford. When we say complex systems we actually mean something pretty precise when we say that. We’re not talking about simply a complicated system.

For example a locomotive train is very complicated but I want to call it complex. When we say complex what we usually mean is that it’s a system that’s made up of many different agents, that are all interacting and there’s some form of emerging behavior. The classic example of complex systems is a flock of birds. You can take each one of those birds and you can tell it to follow just 2 simple rules. For example stay as close to my neighbor as possible but don’t run into my neighbor.
[00:48:00] If you tell all the birds to do that and then you put them together in a big flock you get these wonderful flocking patterns that, are incredibly complicated and they can avoid predators incredibly well. You’d never be able to predict or understand any of that emergent macro behavior, by simply looking at the individual birds. This is really the essence of putting all these different agents together and watching them interact. That’s what complexity study is. Complex systems are all around us. Economies are complex.

There’s a lot of really interesting research that’s been done on chaos and [freckles 00:48:17] and cellular or [tomoda 00:48:20] and a tonne of other really interesting topics.

Bill:
[00:49:00] What I find interesting is the interdisciplinary approach. If you’re really deep on your domain and then you have someone that’s deep on molecular biology, and another person that studies in another discipline, what I found it was interesting when I was reading about this group is that, what’s the value of having multiple disciplines together? The example that was on the side was that you have the coordinated behavior or ant colonies where no ants are in charge.

How could that possibly have to do with the production of thoughts in the brain? How both of these phenomena may relate to ancient notions and spirit and the soul? Those are 3 very different areas. What would your complexity group try to approach with that particular example?

Marshall: We’re a huge proponent of interdisciplinary teams. One of the reasons for this is that in complex systems especially what we see are these phenomena, that have universality. Universality basically is that you find the same essence or the same property in many different seemingly unrelated systems. Earlier in the podcast we were talking about power laws and this basically mathematical relationship, of how often something happens versus its size.

[00:50:00] What’s really interesting about power laws is that there are all different seemingly unrelated systems. If you look at the number of Twitter followers people have it turns out that that follow is a power law distribution. If you look at the income distribution in the United States same thing, it’s a power law distribution. You can look at the population of cities. You can look at the lengths of different segments on a tree of how often a tree branches. That also follows a power law.
[00:51:00] The lungs in humans and all mammals also follow a branching pattern that follows a power law distribution. What we see is that there’s this same phenomena that we see in many different unrelated systems. It’s really useful from the interdisciplinary standpoint is to get a lot of different people together. Perhaps if in cyber security we’re really stuck on a certain problem maybe we can bring in a biologist who says, “You know what we had a very similar problem when we were thinking about cancer and this is how we formulated that problem.

This is the solution that we came up with.” Then we can take a look to see if any of those solutions might be applicable to the cyber domain.

Bill: It’s really interesting because I had a guest named Michael Michalko who was a creativity expert. He’s written several books. One of the sections of the book was bringing in [despaired 00:51:23] domains, when you’re trying to solve a problem. Don't solve it from the part of your brain that is related to that discipline. Actually say how would another system solve this problem. I just found it interesting that you were going down that path because that was almost again another discipline.

A creativity expert was actually talking about this as well. It’s interesting you and I are having this conversation about potentially solving problems using other disciplines.

Marshall: Yeah absolutely.

[00:52:00]
Bill:
Listen Marshall this has been a fascinating conversation and I really want to congratulate you on your work. It’s really profoundly informative and practical for the CSO's and the CIO's that are from the small, medium, and large businesses, in United States and across the world. I appreciate you for your depth of insight and research in this area.

Marshall: Thanks I appreciate it. Thanks for having me on.

Bill: I look forward to when you get to your thesis not your thesis your-

Marshall: Dissertation.

Bill: Your dissertation when you have that published I’d love to have you on for a round 2.

Marshall: Sounds good.

Bill: We’ll link up all of your presentations and links to the books you recommended and also so people can get in touch with you. What's the best way? Through your Stanford website if they want to reach out to you?

Marshall:
[00:53:00] Yeah absolutely. Feel free to reach out to me. The best email to reach me at is mtypers@stanford.eu. Also you can visit my website. The easiest way to find it is to just google my name and Stanford. We have a couple of resources up there and papers that we have published and a couple other documents, that might be useful to some people in the IT field.

Bill: Thanks again Marshall. Have a great rest of the day and look forward to having you on in the future.

Marshall: Thanks Bill.