EPIC2019

Ethnographic Praxis in Industry Conference Proceedings

Bringing the Security Analyst into the Loop
From Human-Computer Interaction to Human-Computer Collaboration

LIZ ROGERS, *IBM Security*

This case study examines how one Artificial Intelligence (AI) security software team made the decision to abandon a core feature of the product – an interactive Knowledge Graph visualization deemed by prospective buyers as "cool," "impressive," and "complex" – in favor of one that its users – security analysts – found easier to use and interpret. Guided by the results of ethnographic and user research, the QRadar Advisor with Watson team created a new knowledge graph (KG) visualization more aligned with how security analysts actually investigate potential security threats than evocative of AI and "the way that the internet works." This new feature will be released in Q1 2020 by IBM and has been adopted as a component in IBM's open-source design system. In addition, it is currently being reviewed by IBM as a patent application submission. The commitment of IBM and the team to replace a foundational AI component with one that better aligns to the mental models and practices of its users represents a victory for users and user-centered design, alike. It took designers and software engineers working with security analysts and leaders to create a KG representation that is valued for more than its role as "eye candy." This case study thus speaks to the power of ethnographic research to embolden product teams in their development of AI applications. Dominant expressions of AI that reinforce the image of AI as autonomous "black box" systems can be resisted, and alternatives that align with the mental models of users proposed. Product teams can create new experiences that recognize the co-dependency of AI software and users, and, in so doing, pave the way for designing more collaborative partnerships between AI software and humans.

INTRODUCTION

In the spring of 2018, some 18 months after its launch, a small team of IBM Security designers began working on QRadar Advisor with Watson – an artificial intelligence (AI)-driven security software application – in hopes that they could improve the product's user experience and increase adoption and usage. Not surprisingly, the design team had lots of questions for the broader product team. What did Advisor do? How did it work? More importantly, how did its intended users – enterprise security analysts – actually use the application, and did they find the information presented meaningful and useful? The answers to these questions, the Advisor design team argued, could not be gleaned from the typical client phone calls but instead warranted an ethnographic study of security workers – analysts and leaders – within the context of their work environment, Security Operation Centers or SOCs, for short.[1] See Figure 1.

Figure 1: Bulletproof Security Operations Center. Source:
http://media.marketwire.com/attachments/201702/72527_bulletproof-SOC-large-tiny.jpg

SOCs are typically staffed by experienced teams of security analysts and engineers, incident responders, and managers who oversee security operations. They tend to be rather imposing, dark spaces filled with security team members in their own workspaces, surrounded by at least two if not three screens. These teams are responsible for protecting company assets from security threats, which they do by monitoring, detecting, investigating, and responding to potential security breaches. Security operations teams uses a range of technologies, software, and security processes to help them collect, monitor, and analyze data for evidence of possible network intrusions. One such software application is QRadar Advisor with Watson (Advisor). Advisor is designed to help analysts focus on the most critical threats to their network, investigate these threats more quickly, and identify possible breaches that weren't identified by other tools.

Building enterprise security software requires deep knowledge of information technology, the software development process, and the cybersecurity industry. While product teams need to understand the practices, experiences, and goals of their intended users, they also need to understand the technology behind the software. This can be particularly challenging for designers and design researchers who don't come from a computer science background. As a result, it is not an unusual for IBM designers and design researchers to spend significant time when starting a project trying to understand what the software they work on is supposed to help users accomplish and how.

The introduction of designers and design researchers to development teams, however, has proved to be just as challenging for software developers and product managers who are not accustomed to being asked to think about their users' "as-is" experience of their product, complete with pain points and opportunities for improvement.

QRadar Advisor with Watson today, by all accounts, is a complicated application: hard to configure properly, difficult to use, and not especially clear in the insights that it provides analysts. Designed and developed by software engineers more intent on making the backend technology work than the providing an intuitive and frictionless user experience, Advisor has encountered resistance from analysts who don't know how to use or interpret core features of the application. In addition, the application is not particularly well integrated into the broader software system in which it is embedded. Analysts can accomplish many of same tasks facilitated by Advisor, although not as quickly or easily.

Given the complexity of the product and uncertainty around how exactly analysts were or weren't using the application, the lead design researcher of the team lobbied for direct access to analysts and their colleagues within their work environment. It was only in observing and talking to security analysts and leaders doing their work within the context of the SOC that she felt she could properly understand how these workers did their job, why they preferred certain tools and resources over others, and their goals in using or purchasing the tools they did.

After first presenting a more technical description of the Advisor application, this paper provides some background on the field of cybersecurity and the hopes and fears associated with AI within it and the world it inhabits. The paper then proceeds to summarize the specific research goals and methods of the project, key findings, and research outcomes. It concludes with a summary of the project.

QRADAR ADVISOR WITH WATSON

QRadar Advisor with Watson is a cloud-based application that is used by security analysts and incident responders to augment the capabilities of QRadar, an industry-leading security information and event management tool (SIEM). Companies employ SIEM solutions to monitor their environment for real-time threats and catch abnormal behavior and possible cyberattacks. QRadar, like other SIEMs, works by collecting and normalizing log and flow data coming from network infrastructure, security devices, and applications and comparing this data to pre-defined rulesets. If the conditions of a rule are met, QRadar generates an "offense" – a grouping of related "events" that have occurred on a network's devices – which serves to alert security operations that a possible breach in security has occurred. These alerts often are the first clue that there may have been unauthorized access and use of enterprise assets. Unfortunately, many of the alerts that are triggered by SIEMs are false alarms, and security analysts spend much time trying to ascertain if the alert is a true or false positive.

QRadar Advisor with Watson is designed to help security analysts quickly reach a decision on what to do next after receiving one of these QRadar alerts. Prominent in marketing materials is Advisor's status as an AI-enabled application. See Figure 2.

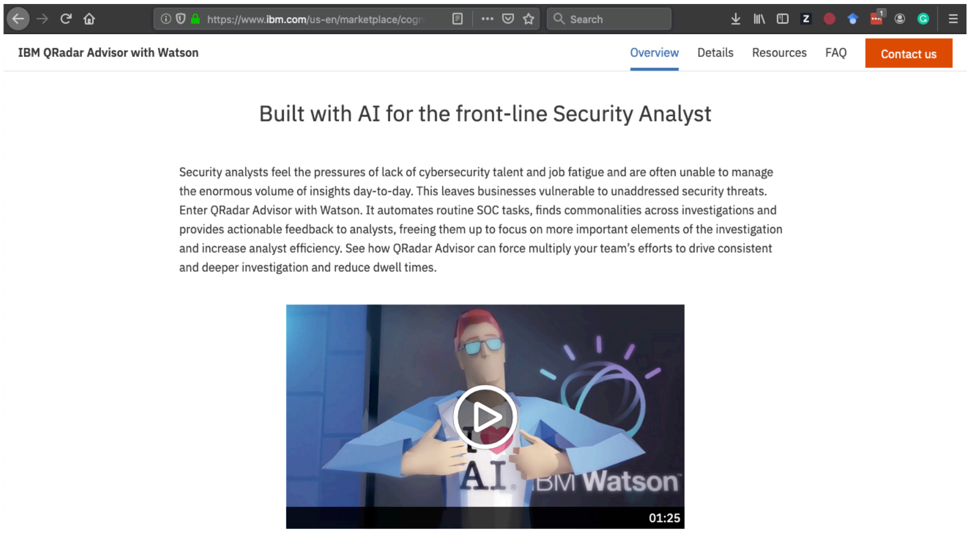

Figure 2: IBM's QRadar Advisor with Watson. Source: https://www.ibm.com/us-en/marketplace/cognitive-security-analytics

Advisor collects internal data from network logs and security devices like firewalls and antivirus devices and correlates this data with external threat intelligence that it has mined from the web. Advisor uses a Natural Language Processing (NLP) model to extract and annotate the external data, which are stored in a knowledge graph (KG). This is the "AI" or "Watson" part of the application. Knowledge graphs are powerful tools that can be used to show all of the entities (nodes) related to a security incident (e.g., internal hosts, servers, users, external hosts, web sites, malicious files, malware, threat actors, etc.) and the relationships (edges) between these entities. Figure 3 depicts an Advisor investigation of a security incident. The result is a comprehensive view of all of the entities involved in the original QRadar offense, along with additional entities in the network that have been identified by Advisor as being potentially affected based on the threat intelligence it mined using the NLP model.

Knowledge graphs, however, can get quite complicated, especially as security incidents can involve hundreds of nodes and edges. See Figure 4 for an example of an Advisor investigation of a complex security incident.

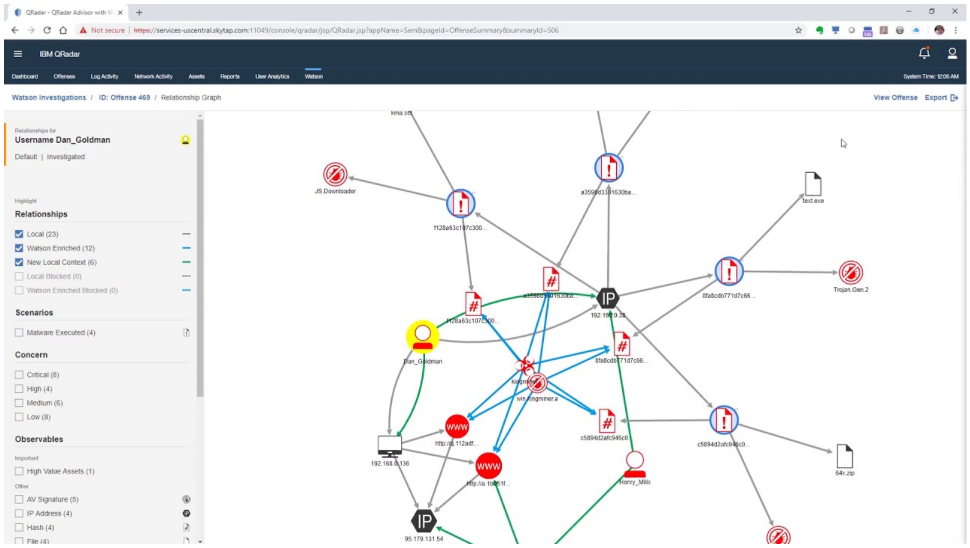

Figure 3: QRadar Advisor with Watson investigation. Source: https://www.youtube.com/watch?v=a5xaY6THvKo

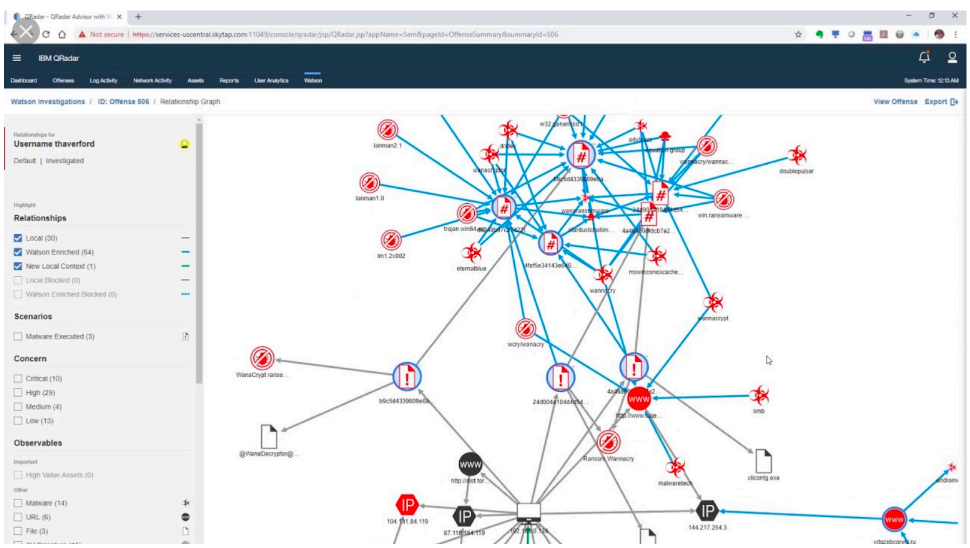

Figure 4: QRadar with Watson Advisor Investigation. Source: https://www.youtube.com/watch?v=NaGpfttxA2s

BACKGROUND

Cybersecurity and AI Technology

In a recent 2019 Capgemini survey of 850 senior executives from 7 industries and ten countries, 69% responded that they would only be able to respond to cyberattacks with the help of Artificial Intelligence (AI). And why shouldn't they think so? AI for cybersecurity has been deemed "the future of cybersecurity" (Forbes 2019). According to at least one company making AI-based security software, AI is "liberating security" from "regular outmoded strategies to one of security as a "science" that brings with it "revolutionary change" (Cylance 2018). There is, of course, another side to the public debate over the impact of AI on the security industry. Customers have voiced disillusion with the over-promising of what AI- and Machine Learning- (ML) based solutions can do. Moreover, cybersecurity experts have warned of the "malicious use of artificial intelligence technologies," based on their prediction that companies will experience new bad actors who are using AI technologies themselves to exploit new enterprise vulnerabilities associated with AI systems (Future of Humanity Institute 2018).

While security experts might see AI as liberating security, AI experts outside of the security community appear to be far less optimistic about the possible effects of AI. For example, based on a 2018 survey of 979 AI experts, Pew Research Center reached the following conclusion: "Networked AI will amplify human effectiveness but also threaten human autonomy, agency and capabilities" (Pew Research 2018: 2). Although some AI experts did recognize possible benefits of AI – e.g., advances in science and humanitarian efforts – on the whole, the experts polled by Pew appear to have far more confidence that the negatives will outweigh the positives. For these skeptics, the adoption of AI technology will result in humans' loss of control over their lives, their jobs, the ability to think for themselves and, the capacity for independent action. (Pew Research 2018). AI technology, according to the study, could lead not only to a rethinking of what it means to be human but also to the "further erosion of traditional sociopolitical structures," "greater economic equality," a divide between digital 'haves' and 'have-nots', and the concentration of power and wealth in the hands of a few big monopolies.

Pervasive Social Meanings of Computing

People have worried about the debilitating effects of new technologies since well before the emergence and popularization of Artificial Intelligence. Computing, in particular, has been a lightning rod for both proponents and critics of the power of technology to transform society and humanity's relationship to nature and the material. Since its introduction, the computer has quickly come to be seen as evidence that routine clerical work could be mechanized and automated – a good thing, confirmation that humans could be freed from repetitive labor and technology was a source of continual growth and prosperity (Prescott 2019).

This vision of computing, like those of previous technological innovations – e.g., steam railways, automobiles, radio and electricity (Pfaffenberger 1988; Moss and Schuutz 2018) – has much to do with Enlightenment ideas of progress and the transformative social potential of technology. This notion – that technological innovation represents human progress and mastery over nature – forms the backbone of a "master narrative of modern culture" (Pfaffenberger 1992). In this master narrative, human history is a unilinear progression over time from simple tools to complex machines. Accordingly, computers are evidence of humanity's increasing technological prowess, control over the natural world, and application of science. They are, in short, a root metaphor for social process in mechanized societies (Ortner 1973).

Not all people have embraced this master narrative, of course, and people seeking to reassert human autonomy and control in the face of mechanization resist and challenge these dominant meanings in numerous ways. For some, resistance comes in the form of introducing new technologies that subvert or invert commonly held meanings of existing technologies. Thus, the invention of the personal home computer can be seen as a strategy to reassert human autonomy and control through the subversion of dominant meanings and images associated with large-scale enterprise computers (Pfaffenberger 1988).

Others undermine this master narrative of technology and progress by subverting dominant themes and meanings attributed to new technologies like AI. Researchers like Moss and Schuur (2018) and boyd and Crawford (2014) have pointed out how the meanings and myths of AI technology and big data have contributed to an understanding of technology as objective, accurate, and truthful, and an understanding of humans as fallible, inefficient, and ripe for machine domination. Other researchers have focused on making people aware of just how dependent machine learning and AI models and algorithms are on humans (see, e.g., Klinger and Svensson 2018; Seaver 2018). As Seaver (2018) has argued, "In practice, there are no unsupervised algorithms. If you cannot see a human in the loop, you just need to look for a bigger loop." Still others have drawn attention to inaccuracies in the master narrative: AI is not objective; there are biases in machine learning models and algorithms.

In exposing taken-for-granted truths about AI technology as myths, these researchers can be seen as authors of a counter-narrative. These counter-narratives do more than just call into question this master narrative, however. They question one of its fundamental precepts: namely, that technology is an external, autonomous force that develops according to its own internal logic. In so doing, these counter-narratives make way for understanding how technologies (and the material) might acquire agency and function as agents in society.

From Humans vs. Machines to Humans + Machines

As AI technology becomes more and more sophisticated, it is hard to imagine not seeing AI artifacts as displaying agency and even autonomy. Even before the popularization of AI technology, however, agency – in particular, the notion of nonhuman or material agency – has been a rich source of discussion and inquiry for a variety of disciplines. Two approaches – one, techno-centric and the other, human-centric – both have been roundly criticized: the first, for its unproblematic assumption that technology "is largely exogenous, homogenous, predictable, and stable, performing as intended and designed across time and place"; and the second, for its minimization of the role of technology itself and its focus on the human side of the relationship (Orlikowski 2007).

In contrast to these approaches, "post-humanist" conceptualizations of the human-material relationship have been proposed that try to avoid the determinism of early concepts and challenge traditional approaches that restrict agency to humans. These alternative concepts bring attention to the way in which humans and technology are inextricably entangled and mutually constitutive in practice. Moreover, they challenge notions of agency proposed by these other approaches. Agency is no longer defined in terms of an essential quality inherent in humans -- a "capacity to act" ala Giddens – but as "the capacity to act" within "entangled networks of sociality/materials" (Orlikowski 2007). Agency is something that occurs rather than something that one has. Both humans and machines thus can be understood to demonstrate agency in the sense of performing actions that have consequences, but both kinds of agency are to be seen as intertwined, not separate (Rose and Jones 2005).

Neff and Nagy (2018) have gone so far as to argue the "symbiotic agency" is a more appropriate expression to capture the dynamic nature of human and technological agency in human-machine communications, in which users simultaneously influence and are being influenced by technological artifacts. Research that has embraced this way of conceptualizing the human-machine relationship recognize people's routines and technology as flexible, especially in relationship to one another: people will change their existing routines when faced with new technological tools and features, just as technological tools and features will be resisted and/or modified – i.e., their material agency will be changed – by people who aren't able to achieve their goals given the current tool or technology (Leonardi 2011). How people work, then, is not determined by the technologies they employ, regardless of how constraining they might be. Instead, people are capable (within existing conditions and materials) of "choosing to do otherwise" in their interaction with technological tools (Orlikowski 2000).

RESEARCH GOALS AND METHODS

At IBM, design researchers need to be scrappy. Getting access to users of IBM products can be particularly challenging, and researchers often do not have the budget to pay for things like recruiting, transcription, and incentives for non-client users. Working for IBM Security adds additional complications. Many of IBM's security clients have mature security operations that have extended teams protecting their systems. Clients can be very reticent to share screens that include real network data or information that reveals how they have set up their security tools for fear of revealing their network vulnerabilities and compromising their security posture. More common than field visits to client Security Operations Centers or even video calls, then, are phone calls attended by members of a client's security operations (which may or may not include people who actually use the product) and interested IBM parties (e.g., technical salespeople, offering managers in charge of the business, engineers, and designers).

There is only so much, however, that can be gathered from such phone calls, and initial calls with Advisor "users", while informative, did not provide the team with a thorough understanding of the processes and tools used by security analysts, the goals they have in using these, and the constraints that they encounter in trying to accomplish these goals. Ethnography, the design team argued, would help them understand how analysts interacted with and made sense of the "data overload" and "noise" that marketing materials referenced.

Thus, in the late summer of 2018, IBM design researchers working on Advisor were permitted to shadow a handful of security analysts and leaders in their workplace. This research occurred in May and June 2018 and included visits to the SOCs of two IBM clients:

one, a large Managed Security Service Provider that uses IBM security solutions to provide security services to more than 500 customers; and the second, a large distributor of manufactured components with a security team of 10 globally-distributed people.

Researchers spent three days at the first of these two SOCs, and one day at the other. While visiting the two SOCs, researchers shadowed six different security analysts and met with one security leader and his direct reports. Visits, with permission, were taped using an audio recorder and transcribed afterward. Researchers did take pictures, although these cannot be shared as they contain client data. Research goals centered on the following three objectives:

- Understand how security analysts currently monitor threats and analyze, diagnose, and triage security incidents and what drive these behaviors;
- Understand how analysts are and are not using Advisor today to help them meet their objectives and why; and
- Identify how the team might improve Advisor so that security analysts can complete their investigations more efficiently.

Findings and recommendations from the ethnographic research were used to fuel an internal workshop that led to the identification of three user goals to guide the design and development of the next major Advisor release. The following user goal drove the reinterpretation of the graph: *An L1 security analyst can view what really happened in their network for a potential incident and complete triage five times more efficiently.*

After the workshop, additional user research was conducted to "validate" user needs associated with each of the identified goals and assess how different design concepts developed by the team did or did not help users achieve the stated goal. Most relevant to this case study are interviews with five security analysts recruited through respondent.io that focused on gathering user feedback on a set of alternative concepts, as well as discussions with eight additional security leaders and analysts from five different Advisor clients regarding the final design concept.

KEY FINDINGS

Competing for Analyst Mindshare

Finding #1: Security analysts are reticent to incorporate new tools into familiar work routines, especially if they trust their existing tools and are effective in using them.

Security analysts have many tools and resources – open source, public, and commercial – at their disposal to help them monitor network traffic for suspicious behavior and activity. Besides QRadar, the research team witnessed analysts using an array of network security devices (e.g., antivirus, firewalls, intrusion detection and intrusion prevention systems), threat intelligence feeds, anomaly detection and user behavior analytics, network access controls, and application-, network-, host- and infrastructure-related log collection. Information overload is a real problem for security analysts, especially because many of these tools and data sources are not well integrated, forcing analysts to manually dig through these sources of data and correlate them.

With all the data they must collate and dig through, security analysts have developed their own practices and strategies, strategies which include the use of popular free tools and data. QRadar Advisor competes with these existing tools and resources in the minds of analysts, and it doesn't always win.

> "I don't know if I really use it [Advisor] that much, because I have so many other tools that I'm looking at on a daily basis." — Security Analyst

The Need for the Human Element

Finding #2: Security analysts rely on their own personal experience and knowledge of their network to assess if an offense is evidence of a breach or a "false positive."

The QRadar offenses investigated by analysts often are complicated, and the tools that they use are imperfect. Prior to starting an investigation, security analysts want to know which offenses to work on first. Offenses are not all equal in how critical they are to an organization, and not all offenses represent an actual security breach. Critical offenses are those that represent great harm to an organization, its reputation and digital assets. They often involve privileged users with system privileges or data access rights that others in the company don't have. Imagine if a phishing attack successfully compromised the Chief Financial Officer's laptop. That would be a critical security incident.

Sometimes offenses are "false positives," however, meaning a breach did not actually occur. There are a number of reasons why false positives happen, including: the rules are not tuned well enough to be able to recognize an action or event as benign, an application does not have access to all of the internal security data that is generated by a large network, and threat intelligence is not nuanced to distinguish URLs that are fine but are hosted on an IP address deemed malicious. As one security analyst told the team:

> "I've had in the past where you guys have flagged legitimate traffic as, you know, malicious, and once I go down to the URL level, and I look at your threat intelligence, you guys have flagged a different site. It's hosted on the same IP, but I get 20 false positive offenses because there's some article about some celebrity hosted on some website in India where it's hosted on the same IP. And we operate in India, I've got staff, they're allowed to read the news, and when they come online, they share the story … and I get a flood of offenses, and I go wild thinking like, 'Oh crap, we're getting like a mass infection event or something.' And it turns out it's not incorrect intel, but intel being incorrectly applied." – Security Analyst

Security analysts believe that there is no solution, powered by AI or not, that can completely know their network like they do. Not surprisingly, then, security analysts are suspicious of claims around automation and of AI omniscience: "Trust but verify" is a mantra the team has heard over and over in working with security analysts. Security analysts recognize that software is imperfect, and they see themselves as filling in the gaps of their security tools by providing the "human element."

> "You have rules that caused the action to fire. In most any kind of programming, you cannot account for all variables. That's why you still have to have the human element to this, because it could be a benign thing between local and local. But it could easily be remote to local or local to remote with the same type of activity." — Security Analyst

Prioritizing Immediate Versus Potential Threats

Finding #3: Security analysts are more focused on protecting their organization's security posture from immediate threats than hunting down potential threats.

In conducting ethnographic research, Advisor researchers discovered that security analysts focus more on identifying "what really happened" during a security incident than "what could have happened." The work of analysts consists of "putting together the trail to determine what happened or caused the issue." Things that "might have happened" or "could have happened but didn't" are simply of secondary importance for them.

> "That's the whole point of the [SIEM] analyst. You have to analyze this data and come up with what's going on. You have to be an archaeologist of IT as you mine the information." — Security Analyst

> "In my field, ultimately it's making sense of a lot of information and trying to glean what caused the incident generally after the fact. It's a lot of firefighting." — Security Analyst

Because analysts are so focused on the highest priority incidents, most of them do not feel that they have the time (or the mandate) to hunt for threats in their network proactively. This prioritization of immediate over potential threats has had a direct impact on security analysts' approach to Advisor and its knowledge graph. At the time of the research, analysts perceived Advisor as a tool for "threat hunters" that "have the time . . . to keep delving."

> "This here [graph] gives the customer … the chance to look at these other IPs because they have time, they have resources, to look at this and further research it. We are dealing with events that are occurring." — Security Analyst

In the eyes of security analysts, their job is different than that of threat hunters": "An analyst's job is purely to look at the security posture, the security stance. Was that a breach? Was there an issue?"

A Confusing Knowledge Graph

Finding #4: Security analysts, especially less experienced analysts, do not know how to interpret the graph and thus do not understand the value it brings to their work.

Spending time in the SOCs, the research team concluded that limited adoption and usage of Advisor was the result of not one but several factors. Unfortunately, not all of these variables could be addressed by the Advisor team. For example, network topologies are often out-of-date, and, as a result, QRadar does not have an accurate or comprehensive view of the entire network. Solutions to this challenge were deemed out of scope for the project. The research team, however, did believe that there was one issue that could be addressed to great effect. Security analysts, the lead researcher argued, did not see value in the graph because the graph was confusing and didn't present information in a way that answered the questions analysts pose in determining the nature and extent of a possible breach.

On the one hand, security analysts' decision not to launch an Advisor investigation can be seen to be the result of their interpretation of how Advisor works and the information it provides.

> "My understanding is that it's an assistant to pull QRadar info in so you don't have to go through all of this QRadar information . . .so with QRadar being pulled in, if you get this message here [in the Insight paragraph of Advisor] saying we found nothing, then you're not clicking on Investigate, it's all working background." – Security Analyst

On the other hand, the research also suggests that analysts are hesitant to use Advisor because of the complexity of the knowledge graph and their difficulty in knowing how to use and interpret the contained information.

Analysts, the research team discovered, want a solution that brings together all of the disparate information they usually have look up manually and presents it in such a way that they can quickly answer the following questions:

- Was a connection made from inside the network (by a computer, a device, an application, etc.) to an IP or URL that is associated with threat actors and attacks, or was it blocked?
- If a connection was made, is it a local-local connection or a local-external connection?
- If a local-external connection was made, what local assets are involved, and are they critical assets (e.g., the computer of the company's Chief Financial Officer)?
- If a local-external connection was made, was malware actually executed by a user?
- What type of attack (e.g., malware; phishing, denial of service) is being used against the network?
- Is this an evolving attack or something that has been contained?

This set of questions determines the workflow of analysts, as seen by one analyst's narration of the information that he was looking for while he was using QRadar to investigate a security incident:

> "Was a connection between the remote host (and malicious observable) and local host made, or was it blocked? If it was blocked, is the system still trying to connect to it (e.g., it's a botnet CnC)? Is the local asset infected? What is the local asset that is in connection with the malicious observable? Who is the user? Was a payload locally executed? If executed, which assets have been compromised, in order of

priority? What has happened over the past seven days? Are new events being added to an offense?" – Security Analyst

In asking these questions, security analysts are attempting to quickly understand the following:

- If a breach has occurred or not
- The source of the breach
- The assets that have been affected and how critical they are
- The kind of attack they are dealing with
- How widespread the attack is

Together these variables allow an analyst to "put together the trail to determine what happened or caused the issue." Very few security analysts the team met could answer the questions listed above with Advisor's current knowledge graph. As a result, they could not quickly come to an understanding of the security incident.

Here, some explanation of how the product team intended security analysts to use the knowledge graph is warranted. For illustration purposes only, Figure 5 depicts an Advisor investigation of a simple security incident.

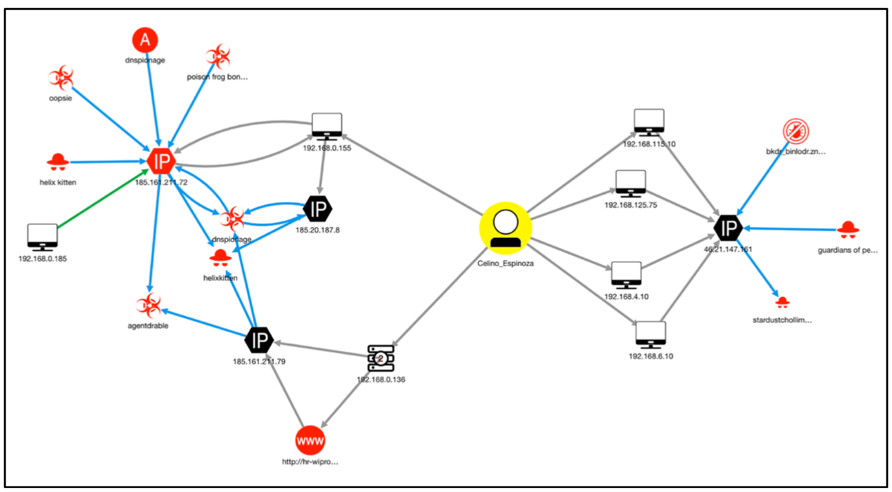

Figure 5: Initial Graph Generated by an Advisor Investigation. Source: SANS Organization 2019

Figure 5 depicts a security incident that can be summarized as follows: six different local assets (the black computer icons) associated with user Celino Espinoza (the yellow circle in the middle) have reached out to three different IP addresses (the black hexagons). In one case, it appears that the user went to a suspicious URL that is hosted on one of these IP addresses. All of the IP addresses in the graph show evidence of connection (and possible compromise) to a whole set of suspicious entities (all of the red icons) like malware, a threat campaign, threat actors, and a virus. Clicking on any of these icons will pull up additional information about that node that can be used by an analyst to understand how critical the threat is. Clicking on any of the lines connecting them (edges) will bring up information about the nature of the relationship between two entities. Hovering over the IP addresses will bring up a geolocation map of where the IP address is registered and physically located.

While the graph provides a lot of useful information, analysts were not confident that it would help them quickly determine if an alert was a true or false positive and what their next steps should be. Analysts specifically mentioned the following as limitations of the current knowledge graph:

- The graph does not clearly indicate the entity that is the source of the offense or attack: i.e., where the attack entered the network.
- The graph does not clearly distinguish between which entities are inside of the network and which ones are outside of the network.
- It is not apparent what was blocked and what wasn't, what was downloaded and executed versus simply downloaded, making it difficult for the analyst to recognize and prioritize immediate threats over potential threats.
- The graph does not clearly indicate which potentially compromised machines are the most valuable, vulnerable, or critical.

Because of these limitations, analysts were often unclear of Advisor's value proposition, regardless of the marketing materials. Was the graph there to help them find the "root cause for an action to fire" and thus save them valuable investigation time? Or was it possible that Advisor was doing the entire investigation of the source offense for them? Was Advisor helping them identify additional indicators of compromise outside of an offense that they would have missed without seeing them on the graph?

Competing Meanings of AI and Advisor

Finding #5: Security leaders and analysts attribute different meanings and goals to Advisor's knowledge graph, resulting in different perceptions of the value of the application.

When presented with a demo of the knowledge graph – say, at conferences or in sales-related talks – security leaders invariably respond positively to it, describing it as "cool," "complex," and "impressive." They can imagine themselves projecting it up on the wall of their SOC or using it in reports for management. Indeed, one of IBM's security executives admitted to the team that potential customers often found the graph the most exciting aspect of the application. Another internal consultant familiar with presenting the application to potential clients called the KG visualization "eye candy" for security leaders. The research team's conversations with security leaders also revealed their admiration for the diagram:

> "It's complex, and it can impress people. You can put it up on a screen and show senior management, and they'll go 'wow!'. . . and it looks like the Internet. It looks complex and impressive." – Security Leader

This insight – that senior management favored the KG visualization much more than analysts did, based on the status and prestige it presumably conferred – was a revelation to the team. Security analysts with little to no experience of the Advisor, however, characterized the existing knowledge graph as "this big spider web" that "displayed too much data in a format that wasn't clear." For them, the knowledge graph is an intimidating artifact that is difficult to interpret and hard to verify.

RESEARCH RECOMMENDATIONS

Despite the different meanings attributed to the knowledge graph, the Advisor team continued to believe in the value of a graphical representation of a security incident, however elusive. Creating such a representation is akin to finding the holy grail for the security industry.

> "If you could get a graphical representation that shows you what you're looking for and at least points you in the right direction, it's worth a million bucks — compared to going through ten thousand rows, trying to find it yourself later, adding filters on filters on filters, trying to figure out what caused it or what happened. Trying to make sense of the data … dividing it as granularly as you can without losing it in the noise." — Security Analyst

Research recommendations based on the ethnographic research did cover the knowledge graph, as well as other opportunities for improving the solution not discussed in this case study. Suggestions for how to improve the knowledge graph included the following:

- Explore new ways in which to organize the information mined by Advisor. Are there different metaphors that can guide the visualization of the graph? How can we align the diagram better to the mental models of analysts?
- Clearly distinguish between "what happened," "what didn't," and "what could have happened" in the knowledge graph — i.e., distinguish between the actual path of attack and any "what if" scenarios.
- Help analysts get started investigating by providing them with a quick, cursory overview of what they are dealing with.
- Allow users to keep digging from within the graph easily.
- Allow users to investigate offenses related by malicious observables, as well as known attack tactics and techniques.
- Identify which potentially compromised machines are the most valuable, vulnerable, or critical.
- Leverage users' strategies to distinguish between legitimate and illegitimate traffic and identify the incident type. Show them which connections were made and which ones were blocked. If connections were blocked, is the local host still trying to call out to the remote IP?

RESEARCH OUTCOMES

These recommendations, along with an "as-is" investigation workflow, were the cornerstone of a 3-day workshop, in which the team identified three main experience objectives for the next major release of Advisor. One of these was: "*An (L1) security analyst can view what really happened in their network for a potential incident and complete triage more efficiently.*" This goal became the north star for the Advisor team working on a new visualization of the graph.

Putting Together the Pieces of the Puzzle

At the forefront of the minds of the two designers tasked with creating a new knowledge graph visualization was the desire to create something that would help analysts "connect the dots" so that they could tell the story of what had happened. Both designers recognized that the previous visualization, while technically correct, was not very consumable nor did it meet the goals the team had for themselves for designing for AI:

> "We're the kids with a messy room when we create products. Something that may seem chaotic or out-of-place to our users doesn't seem so crazy because we created it. We live in this room, in our products. But we need to create something consumable, constructive, and structured when it comes to data visualization and bringing forward explainability and transparency from artificial intelligence." – Advisor Designer

Designers use metaphors to explain the new and unfamiliar in terms that people – users – understand. If the current visualization of the Advisor knowledge graph brings to mind the complexity of the Internet and the "black box" nature of AI, what then is an appropriate metaphor for a new visualization, wondered the designers.

After much experimentation, Advisor designers landed on a metaphor closer to how security professionals themselves explain their process and what it is that they do — a puzzle. Puzzles are composed of lots of pieces, some of which fit together, others that don't, and still others that might be missing. Their job, the designers explained, was to present analysts with all of the pieces of the puzzle that were available (e.g., the rule that triggered the offense, user and asset information, threat intelligence, malicious observables) and let analysts "fill in the empty gaps."

Using this metaphor, Advisor designers produced several different concepts, one of which featured the use of four "swim lanes." See Figure 6. This visualization of knowledge graph data addresses the primary reason why so many security analysts using knowledge graphs find them so very difficult to interpret, namely the absence of a structured flow through the nodes and edges. With traditional visual representations of a security incident knowledge graph, there really is no easy way to follow the path from the source of the incident to the possible threat, due to the many interrelated branches.

In contrast to existing visualization, this new way of visualizing a knowledge graph reduces complexity by clustering related entities together. Related entities that can be clustered together are determined not only by the type of the entities, but also by the threats impacting them. The new graph representation also provides an easy-to-follow path starting from the source of the security incident – typically a user or an internal asset or an external entity – and leading to the threat that allows the security analyst to quickly identify how the security breach proceeded through their network. And, finally, it reduces the clutter of the old diagram by allowing security analysts to selectively expand clusters they would like to see more details on.

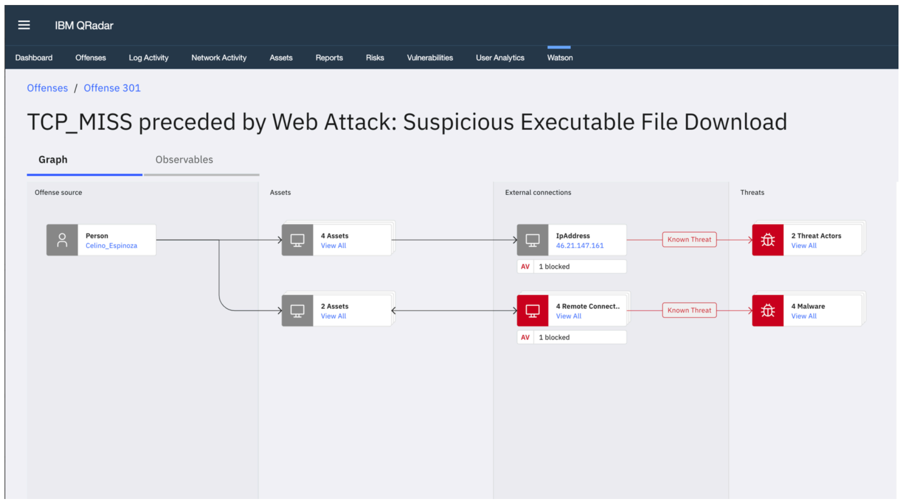

Figure 6: Proposed Knowledge Graph Visualization. Source: IBM QRadar Advisor with Watson product team.

In effect, this new diagram quickly provides analysts with the answers to their questions by mimicking their workflow and aligning with their mental model of how attacks work. The diagram makes clear what the source of the offense and attack is and where the analyst should start the investigation. Also made explicit are the internal assets that are involved in the security incident. The diagram also identifies any external connections that were made to any suspicious or malicious URLs or IP addresses, and clearly calls out if network security devices did or didn't block the threat. Payload information is available from within the diagram, as is additional information about all of the entities and their relationships to each other. Lastly, the type of threat and its success or failure in compromising the network is clearly communicated.

With this new visualization, the Advisor team provides analysts with all the puzzle pieces they need to make a quick assessment if an offense represents a false positive or a real threat.

RESEARCH IMPACT

After the ethnographic research and workshop, the Advisor team worked closely together with security leaders and analysts to develop a KG visualization that met the agreed upon goal of *"an analyst can view what really happened in their network for a potential incident and complete triage more efficiently*." Interestingly, both security analysts and leaders appreciate the new diagram and for similar reasons.

> "The new concept would absolutely be easier to determine if it is a false positive or if something needs to be looked into more or escalated. It's much easier for us to see the flow of what was going on." — Security Analyst
>
> "It has the [data] structure, the involved information, and clear definitions of the types of connections and assets." — Security Analyst
> "Honestly, I really appreciate the way the information is organized on this graph. It's A LOT cleaner. We have had many offenses when the investigation will have several hundred IPs on it, and it's just almost impossible to easily glean important information out of those. There's just so much clutter on them." — Security Leader

It is true that some security leaders asked if it was possible for the Advisor team to support both diagrams. When told "no," however, security leaders opted for the new diagram, undoubtedly in part because of their own background as analysts.

A new version QRadar Advisor with Watson complete with the new KG visualization will be released in Q1 2020 by IBM. Time will tell if the new graph diagram will increase usage and sales, but the team (including upper management) remains confident that they made the right choice. This certainty is, in large part, due to the research that drove the decision to work on a new knowledge graph visualization and research that validated the preference for the new diagram.

The design team's work on Advisor has also had an impact on how teams are approaching designing for AI at IBM. The design team regularly consults with teams across the business on how it arrived at the user-centered goals that drove the development of a new AI-powered experience. In addition, the graph has been adopted as a component in IBM's open-source design system and is currently being reviewed by IBM as a patent application submission.

CONCLUSION

People, in general, are conflicted about AI. According to one dominant narrative, humans will likely experience a future made more productive and efficient by Artificial Intelligence. Counter-narratives, however, predict a different kind of future – one in which humans become less autonomous and in control of their lives and are incapable of making decisions and taking action independent of AI tools and technologies. The customers and users of QRadar Advisor with Watson are no different. They believe in the power of AI to advise them of what they don't know and what they should do, but they also question the ability of – and their desire for – a tool, any tool, to replace them, the human element.

Like our users, AI enterprise solution product teams are conflicted. They believe in the power of the AI products they are designing to benefit the lives of users, yet they also recognize that they are developing products whose goal is to reduce the need for human effort (or what are wistfully thought of as "lower order" tasks and skills).

Humans are – and always will be -- a necessary part of the equation. Humans are not just consumers but active producers of the insights that AI models produce. Humans are the agents that create the interfaces and visualizations that people use to interact with AI models and AI-generated insights.

In exploring how the Advisor team came to the decision to replace one KG visualization with another, this case study demonstrates just how entangled humans and technology can be. It also suggests that AI agency and autonomy are less of a threat to humans and human agency than certain parties would suggest. Could it be that Artificial Intelligence is really more of a neutral force whose exact influence shapes and is shaped by humans engaged with it.

That change will occur with the introduction, adoption, and adaptation of new technologies is certain. The exact nature of this change is not, however. In challenging the way in which AI-powered insights are represented to analysts and proposing a solution that better aligns with their own mental models, the Advisor team undermined the notion that humans have no role in the future or expression of AI. It took people – designers, engineers, offering managers, security analysts, and security leaders committed to developing a product that users could use and get value from – to find a way to present the information in way that was consumable and, in the process, reveal the co-constitutive nature and required human element of AI. In so doing, they call attention to the ways in individuals can challenge the trope of AI as the harbinger of a future in which individuals are made more productive yet less autonomous.

Recognizing humans and nonhumans as partners in a symbiotic relationship challenges the concept of "human-computer interaction." Designing from a shared agency perspective means that product teams must consider the interdependence of humans and nonhuman actors and design for two entities. As Farooq and Grudin (2016: 32) argue, "The essence of a good partnership is understanding why the other person acts as they do. Understanding the intricate dance of a person with a software agent requires longitudinal or ethnographic approaches to produce realistic scenarios and personas."

Liz Rogers is a design research practice lead with IBM Security. After receiving a PhD in cultural anthropology from the University of Wisconsin – Madison, she entered the world of product innovation and never looked back. She has over 18 years of experience working in the design industry, helping teams design compelling products and experiences, based on a deep understanding of user needs, motivations, and behaviors.

NOTES

Acknowledgments – I want to express my gratitude for the support, encouragement, and thoughtful engagement provided by Patti Sunderland, the curator for this paper. Without her, this paper simply would not exist. I also want to thank Terra Banal and Andi Lozano, the two designers on the Advisor team. Both are amazing designers and people. Without them, there would be no new knowledge graph visualization, nor would the research have been as successful as it was. Lastly, I'd like to thank the entire Advisor team, as well as the security analysts and leaders who helped us design the new knowledge graph visualization. I can't wait to see what happens.

1. In smaller security organizations, it is not uncommon for one individual to cover multiple roles, including security leader, security analyst, incident responder, and threat intelligence analyst. Larger, more mature security teams typically distinguish between these roles with differing degrees of granularity. Each of these roles can be identified in multiple ways. A simple search using a website like Indeed.com brings up multiple ways to identify the people who take on the responsibilities and tasks associated with "security leaders" – e.g., creating, implementing, and overseeing the policies,

procedures, and programs designed to limit risk, comply with regulations, and protect the company's assets from both internal and external threats – like Chief Security Officer, VP of Security and Risk Manager, and IT Risk Management Director. Similarly, people whose top jobs to be done include protecting company assets against tools of attack and attacker, detecting the occurrence of cybersecurity events, and investigating the activities and presence of attackers include people working as SOC Analysts, Information Security Analysts, and Security Engineers. For the sake of clarity and simplicity, in this paper, individuals who perform similar tasks and have the same goals, pain points, and needs in performing these tasks are all referred to by a common title, in this case "security analyst" or "security leader."

REFERENCES CITED

boyd, dana and Kate Crawford
2012 Critical Questions for Big Data, Information, Communication and Society. DOI: 10.1080/1369118X.2012.678878

Capgemini Research Institute
2019 AI in Cybersecurity Executive Survey. https://www.capgemini.com/wp-content/uploads/2019/07/AI-in-Cybersecurity_Report_20190711_V06.pdf

Cylance
2018 How Artificial Intelligence Is Liberating Security. https://www.blackhat.com/sponsor-posts/06252018.html. Accessed August 27, 2019.

Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Center for a New American Security Electronic Frontier Foundation, and OpenAI
2018 The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.

Klinger, Ulrike and Jakob Svensson
2018 The End of Media Logics? On Algorithms and Agency. *New Media & Society* 2018 (12).

Leonardi, Paul
2011 When Flexible Routines Meet Flexible Technologies: Affordance, Constraint, and the Imbrication of Human and Material Agencies. *MIS Quarterly* 35 (1): 147-167.

Moss, Emanuel and Friederike Schuur
2018 How Modes of Myth-Making Affect the Particulars of DS/ML Adoption in Industry. *Ethnographic Praxis in Industry Conference Proceedings* 2018: 264-280.

Neff, Gina & Nagy, Peter.
2016 Talking to Bots: Symbiotic Agency and Case of Tay. *International Journal of Communication* 10: 4915-4931.
2018 Agency in the Digital Age: Using Symbiotic Agency to Explain Human–Technology Interaction. DOI: 10.4324/9781315202082-8.

Orlikowski, Wanda J.
2007 Sociomaterial Practices: Exploring Technology at Work. *Organization Studies* 28 (09): 1435 – 1448. DOI: 10.1177/0170840607081138

Ortner, Sherry
1973 On Key Symbols. *American Anthropologist.* New Series. 75 (5): 1338 – 1346.

Pew Research Center
2018 Artificial Intelligence and the Future of Humans.

Pfaffenberger, Bryan
1988 The Social Meaning of the Personal Computer: or, Why the Personal Computer Revolution Was No Revolution. *Anthropological Quarterly* 61 (1): 39 – 47.
1992 Social Anthropology of Technology. *Annual Review of Anthropology* 21: 491-516.

Prescott, Andrew
2019 What the Enlightenment Got Wrong about Computers. https://dingdingding.org/issue-2/what-the-enlightenment-got-wrong-about-computers/

Rose, Jeremy and Matthew Jones
2005 The Double Dance of Agency: A Socio-Theoretic Account of How Machines and Humans Interact. *Systems, Signs & Actions* 1 (1): 19-37.

SANS Institute
2019 Investigating Like Sherlock: A SANS Review of QRadar Advisor with Watson

Seaver, Nick
2017 Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems. *Big Data & Society*. July- December 2017: 1-12.

2018 What Should an Anthropology of Algorithms Do? *Cultural Anthropology* 33(3): 375.