Logan Cox

Information Security Office – University of Utah

The official statement of what the Information Security Office (ISO) team does is “To help ensure that the University of Utah is exposed only to acceptable risks within its educational, research, patient care, and community outreach missions.” I like to view it as making sure that the University of Utah is able to run its business processes safely and effectively. I work for the Information Security Office at the University of Utah as a Cyber Security Event Analyst. I initially believed that if I picked the same organization that I worked for then it would be cheating. I then realized not only can I hopefully give a more in depth explanation of how the data is used in the organization, I can also dig deeper than my current knowledge and gain more insight on our operations. Because the Information Security Office covers the entirety of the University of Utah campus and hospital, I will keep this to only the data in which we handle directly - the data that we deal with every day and have direct use and/or have control over. In this paper I will talk about the data that we collect, use, analyze, store, and manage within the operation of the organization.

Data Management: The U collects credit card information, patient health files, student records, and other various data information because it’s what the University of Utah needs to do in order to run its business processes. The data storage specifically within the Information Security Office is much smaller than the University of Utah as a whole. It’s mainly data to support the operation on the long range of data that the U collects, or in other words we collect metadata. There’s no intrinsic business value to the data that we collect, other than the fact

that it allows the other types of data and process to exist and run effectively. The data sources we collect are primarily log information, network flows, network traffic, and forensic information.

The log information is a wide variety of information that sums up network traffic or network activity. An example of this would be Payment Card information that is collected when you swipe to buy food or an item on campus. We collect a summary of the what card was used, where the payment went to, what time the payment was made, and other important facts about how the data was transferred. The reason we collect something like that is so that if we see an anomaly in where the information went or how it was used we can look into it and try to prevent fraud. This log information can be stored from 30 days to a couple of years depending on how sensitive the information is. If it's something such as credit card numbers we don't want to collect them and store them for a long time because of obvious reasons such as prevention data loss that would result in fraud. We only collect the ones that show anomalies, and once we investigate and handle the case then we get rid of them. Logs such as Wi-Fi leases that we give to people so they can use UConnect are stored for longer. It's not sensitive information and we might need it later to see if we need to investigate a certain user by looking at the traffic they created during the time they had their lease.

Network flows are similar to any other log, except that they are specific to network traffic. They're a basic summary of network traffic that has happened. When two devices (example: a computer and a phone) communicate they send packets to each other. Inside this

packet is a lot of information such as specific requests, headers, IP addresses, protocols, and other network information. This information can be summed up within a network flow. The network flow can show you what protocol you used, the IPs that communicated, how big the message was, and when the communication happened. This allow us to quickly get a brief overview of traffic that is happening, but not get too deep into the actual packets and their content. On the other hand, you have pure network traffic which is completely raw traffic where you can see exactly what happened between the two devices. Network traffic is very dense so it takes a lot of space to store it and search through it. The Information Security office collects both, but only collects most network traffic for 30 days.

The last piece of data that we collect is forensics from devices we bring in to our office. When we find malware on a device such as a computer we ask if the computer has any sensitive information on it, and if it does then we bring it into our office for forensics. We don't want malware to steal any sensitive information on the computer. During the forensics process we make multiple copies of the hard drive so we can do forensics on the computer without manipulating anything on the original drive. If the forensic team finds any sensitive information or any information that's illegal to have then we keep that data for federal investigation. This data can be kept for a month or even years, depending on how long the investigation lasts.

Data Collection: The only data collection that we do and store specifically is the network flows. The Network Flows are collected in real time by a server that runs it through some algorithms that analyze and simply the traffic. All the other data is collected by other University

of Utah Information Technology teams. We can look at, gather, and/or borrow data from other UIT teams if we deem necessary. A lot of the work that the ISO team does is monitoring data. Data collection is not necessary most of the time because we don't ever need to analyze the data past network traffic.

Data Use: The use of data at the Information Security Office is the key to its success. We use data to analyze the Payment Card Industry, implement Data Loss Prevention, find malware on employee devices, prevent malicious attacks, follow federal guidelines, and protect the University.

One of the main uses of data we us is Data Loss Prevention. When anyone with a @utah.edu email address sends an e-mail it is scanned for any potential sensitive information. This information can vary from Patient Health Information, Credit Card Information, FERPA, University Financial Information, and other information deemed sensitive by the U. The reason each e-mail is scanned is because we don't want any sensitive information to be leaked or stolen. Let's say someone is sending social security numbers with first and last names to their personal e-mail to use for work later at home for patient health reasons. Doing work from home might not sound like a bad idea at first, but if you think of the consequences it could have it definitely ends up being much more serious. If that information was intersected by a bad person, or if the computer that the information ended up being stored on got a virus that stole the information we quickly see the many ways the sensitive information could be compromised. So in order to avoid that we ask that work pertaining to any sensitive

information is only kept on work computers and work e-mails. We also have a way of keeping the information more secure by adding “PHI” in the subject title, which makes the e-mail sent in a secure format that the end user can still access easily. This is mainly used if sensitive information needs to be sent to other outside organizations that the University partners with.

Another use of data is monitoring the Payment Card Industry that is used on campus. Because we can have hundreds of transactions happen on campus through our network at once, we monitor it in order to make sure fraud isn’t prevalent within our transactions. Each machine that uses PCI has to go through much more rigorous standards than any other computer. It has to be kept up to date on security standards, but more importantly it has to be set up so we can monitor the data that is being processed within it. That is so when we see an anomaly or a transaction that is being sent an irregular place, we can look into the event and see what the mistake might have been. If we didn’t monitor that traffic someone could find the device on the internet and quickly try to intercept all the financial information coming from it.

What I believe is the most important part about data use in our office is network traffic. There can be millions of packets transferring through our network at once, so how do we use all that data to do our job? The simple explanation is that we let certain tools find a lot of the initial cases we need to look into. We have tools (they will be explained in detail later) that show us anomalies in traffic that it suspects to be malicious traffic. We then look into the traffic and evaluate whether or not the tool was correct. We can also just look up raw traffic if we want to see what a device has been doing on our network. We don’t care what the traffic is as

long as it's not showing malicious intent. Because we have thousands of student, staff, and visitors at the campus and hospital all the time there really isn't traffic that we can just outright block. There's such a wide variety of things the University does so it's hard to identify normal patterns of malicious traffic.

Some of the most important data tools that we use in our job are Qradar, Splunk, Snort, FireEye, and Symantec DLP. Qradar is what gives us an interface of the network traffic. We can search specific Internet Protocol addresses, ports, sizes, and times within our network traffic. You can see almost everything within the traffic information, except some packet content. Splunk manages logs that we get in order to search through them efficiently, it also can analyze them and alert us if it sees anything that we would want to look further into. Snort and FireEye are both tools that examine network traffic for malicious activity and alerts us of that activity. Symantec DLP is what monitors the e-mails to make sure that sensitive information isn't stolen.

Data Automation and analysis: The automation within security is pretty cut and dry for the most part. The automation within our department can be explained thoroughly and walked through step by step with objective data. One example of this is a copyright script I wrote a couple of weeks ago. What it does is search through out entire inbox looking for any emails within the past week that were sent to us by major corporations that show someone downloading their content illegally on our IP addresses. I then take those e-mails and find the IP and time that the information was downloaded. Then I look into the wireless logs to find out who was on that IP at the time when the download happened. An e-mail is sent to every

individual that I find that says they have to sign a form agreeing to no longer download illegal content. Once that form is signed and given to us we let them back onto the internet. Each part of the process is very objective and can be automated fairly easy. This saves us a lot of hours every week and can be easily explained to anyone wanting to know the process. Most of the things we automate are extremely similar to the copyright situation, so we don't have too many moral dilemmas or discrepancies when automating.

One part of our process that we can't explain very well is when we get malware callbacks from a tool we use called FireEye. It looks at and analyzes our network traffic and then makes an event called a "Malware Callback" that signifies that it believes the IP it's calling back on is compromised. We then analyze the event to try and find out if its guess was accurate, then usually look for a large amount of spur-attic traffic to an IP that is known to be malicious. We can usually see if a bad file was downloaded or if it is sending information that it shouldn't be sending, this it almost a guarantee of a compromise. Usually the tool (FireEye) is right, so although we don't know how it does what it does we do agree that it is right about 99% of the time.

Most of the automation that we do doesn't have a big enough affect such as a life or death situation. The only instance in which we could harm someone in our current tasks is if we were to automate password resets on doctor's accounts. If a doctor's account gets compromised and they were trying to prescribe medication or log onto their account during

surgery it could have an impact on a patient's life. We actually almost got to that point without taking into considering those consequences.

There are instances in which we analyze business decisions and the security risk that they could make. If a department wants to purpose a new a plan to work with a company outside of the University network, we have to evaluate if it has a security risk and see the pros and cons of it. We have algorithms that we use to evaluate that risk and they spit out a risk score. This number is then taken by our Chief Information Security Officer and then he uses that number to help make his decisions. This might not be life or death for someone, but it definitely can have an effect on whether or not a department can make an important decision.

I interviewed Corey Roach (Interim Chief Information Security Office) about the future of security and what decisions he could see being automated. He said he could see network traffic being analyzed by a machine such as Watson (IBM's smart computer), because we rely heavily on products such as IBM's Qradar. It would be easy for IBM to implement a machine learning algorithm in the product and have it start looking through the traffic and make decisions to stop malicious activity. Corey is an extremely analytic person and so to him it makes sense to allow these things to happen. The machine is making very objective decisions based off data that is extremely raw and unbiased. It's hard to distort raw network data. A lot of decisions in security are overlooked by humans, but he says that machine learning definitely could solve a lot of the problems we currently have. He looks forward to seeing the impact it makes in the Cyber Security world.

As you can see the use, management, collection, and decisions being made within the Information Security Office are fairly transparent and objective. The future within Cyber Security could get less transparent with the introduction of machine learning, but the implications of it could be very well worth it. If we could “let loose” some algorithms on our network that found and took care of malicious activity on our network, we would create a more secure and safe place for patients, students, researchers, and visitors.

Logan Cox

Sources:

Corey Roach (Interim Chief Information Security Officer)

http://it.utah.edu/departments/iso/