USENIX Security and AI Networking Conference

ScAINet 2018

ALEATHA PARKER-WOOD

Aleatha Parker-Wood is a Researcher/Manager in the Center for Advanced Machine Learning at Symantec and leads a research team focused on protecting users and their data through advanced machine learning. She received a PhD in computer science from University of California, Santa Cruz, for her work on scientific data management. Her work currently focuses on differential privacy for ML and using deep learning for code analysis, with previous work in file system search, forensics, and AI for Go. She has authored several books and articles as well as numerous patent filings, and most recently served as research co-chair of MSST 2017 and on the PC of ScAINet18. Aleatha_ParkerWood@symantec.com

The USENIX Security and AI Networking conference is a one-day invited talk symposium new in 2018, with Symantec as founding sponsor. It aims to bridge the academic and industry communities in the nascent area of security machine learning and artificial intelligence (AI) and provides a complementary venue to peer-reviewed research conferences and workshops such as AISec and the IEEE S&P Deep Learning Workshop. In the spirit of bridging the two worlds, it was co-chaired by an academic, Polo Chau of the Georgia Institute of Technology, and an industry research leader, Andrew B. Gardner, Head of AI/ML and the Center for Advanced Machine Learning (CAML) at Symantec. It was held in Atlanta, GA, on May 11th, with 122 attendees from many major security companies, as well as students and faculty from Georgia Tech, Emory, UC Berkeley, and more. Audience participation was lively, and there was a parallel discussion track on Twitter at the #ScAINet18 hashtag.

In his opening remarks, Andrew Gardner said that it's an exciting time to work at the intersection of Security and AI/ML but that the challenges faced are significant. Security is characterized by adversarial rare events. The data sets are complex, noisy, heavily imbalanced, and, for the most part, private. Unlike colleagues working on computer vision and other computational perception tasks, this discipline still struggles with the basic representations required for learning on programs, graph dynamics, and the unique event streams of security. He went on to note that "as communities, we have tended to work apart. It's my hope that with greater open and collaborative interaction we can define and frame the next generation of grand problems to focus on, in the same way that self-driving cars have led to huge leaps forward in vision."

The first talk of the day was given by Elie Bursztein of Google, who spoke on abuse detection at scale, and talked about the unique challenges faced by security AI. For example, training data for security becomes obsolete quickly. A cat today is much like a cat from a hundred years ago, but a phishing email is constantly evolving. He also noted that context is critical. Two best friends might say, "I'm going to kill you!" while playing a video game, and it will no doubt be benign, whereas the same phrase in a public argument between strangers at a bar might be a huge problem. The model must account for culture, context, and setting to be accurate. Security ML must balance error costs thoughtfully. An account take-over is very dangerous, for instance, so you might choose to err on the side of false positives, locking people out and offering an extensive manual review process to restore access. He suggested relying on humans to adjudicate the long tail of hard cases wherever possible. Finally, security AI has live adversaries. He suggested limiting the amount of feedback you give to attackers in order to make the attack harder to improve, a theme that would later be reprised by David Freeman of Facebook. Last but not least, he noted that if you have a user feedback mechanism, it can and will be weaponized against you. He advised against blindly trusting feedback and emphasized putting feedback into context, filtering, and rate limiting it. Elie's

talk was delivered as a video recording, so unfortunately there was no audience discussion, but he encouraged watchers to tweet any questions at him.

Next, Jason Polakis of the University of Illinois at Chicago discussed fighting CAPTCHA bots. The evolution of AI has made distinguishing bots from real people increasingly difficult, and impersonation is both easy and cost effective. Most of the tasks that we rely on for CAPTCHAs, such as reading distorted text or recognizing named objects in pictures, are tasks that can now be done with human-level accuracy, using free or inexpensive cloud APIs. He demonstrated how an attacker can use word2vec in combination with Google's image recognition APIs to break image recognition CAPTCHAs at 66.6% success per attempt. Adversarial techniques are not yet defeating off-the-shelf image recognition, so those will not prevent bots. The net result is that CAPTCHAs, in order to defeat bots, are increasingly difficult for human users and pose a huge tax on productivity. He suggests that these techniques will need to be replaced in the near future.

David Freeman from Facebook gave a talk on practical techniques for fighting abuse at scale. In particular, he focused on how to bootstrap labeling from a small data set of ground-truth labels. He pointed out that users are both unreliable and too busy to do all your labeling for you, and that a spam label may just mean "I don't want to see this." But if you use those two sets of labels together, create new features independent of them, and avoid feedback loops, you can get much more reliable predictions. To avoid feedback loops, he reminded the audience that you can't just A/B test new security models, because independence assumptions are violated. If you test on a small set and then deploy to everyone, you cannot be sure whether the adversary gave up or iterated to avoid your classifier in the meantime. Instead, he suggested running in shadow mode to not help the spammers evolve, focusing on the spammer's motives instead of the content, as well as using data they don't control, like the social graph.

Sudhamsh Reddy from Kayak gave a talk on the various types of e-commerce bots, both benign (search engines) and malicious (DDoS, content scrapers, click bots, inventory lock-up bots, etc.). He described how simple volume-based metrics, for example, were effective at detecting the majority of bots seen by Kayak, and how using cascading classifiers, from least to most expensive, allowed them to constrain their computation costs. They save costly techniques such as activity-based analysis for low confidence samples and filter the majority into good or bad using lightweight classifiers.

Alejandro Borgia from Symantec discussed the lifecycle of an advanced persistent threat and how to automate the process of doing attack forensics and attribution. Symantec has gone from a highly manual process to a process that still uses analysts but augments them to give them superpowers. Part of that starts with the attack graph, a giant pile of hay to let them find the needle they are looking for. The attack graph contains information about files, machines, locations, and more. They sift the data to learn generalities about attacks, and then look for clusters of similar events. Rather than looking at one enterprise or event, they look across a wide variety of enterprises and events to learn these attack patterns. He mentioned that Symantec had used this framework to discover Dragonfly 2.0, an advanced threat targeting the energy sector, much faster than they would previously have been able to uncover it.

Yogesh Roy of Microsoft offered a talk on finding suspicious user logins in Azure Cloud. They pool users using similarity and use random walks on user locations. Similar users log in from similar locations, and speed of travel can be used to give a reachability score. The analytics aren't that complex in theory, but in practice, it's hard to do at scale in real time. They use Redis as a cache to partition and store model parameters and behavioral data. They have built a graph of activities across many services—with 22.5M nodes, 46M edges, and 245M security attributes—and use that to model probabilities of attack chains ("kill chain connectivity"). They make an inventory of known attack patterns, match their occurrence in the graph, and then use the rest of sub-graph for context, using the kill chain as a basic probability model to constrain the edges and build out connections using stochastic processes. A compute connectivity score is arrived at using the random walk graph. Finally, they use random forests to classify sub-graphs into scenarios. As a final interesting note, he pointed out that anomalous behavior without attack indicators seems to correlate with insider attackers. An audience member asked how similarity was computed, and Roy said it was entirely based on access patterns and their metadata. Additionally, people had several concerns around geolocation in IPv6, which Roy confessed was an open problem for them.

Le Song from Georgia Tech gave a talk on embedding spaces for graphs. Structure2vec addresses a fundamental problem in graphs—designing features for graphs based directly on data. It leverages strengths of graphical models and deep learning together, using an iterative update algorithm parameterized similarly to a neural network to create an embedding space. First it does an unsupervised pass using the features of each neighbor, pooling, and non-linear updates. Then stronger parameters can be learned downstream using supervision. He gave some examples of how to use structure2vec, including comparing code through control flow graphs and using temporal graph features to find fraudulent accounts. The audience had questions about how the update worked and whether it was unsupervised or supervised. Le explained that the training had both a supervised and unsupervised phase, where the unsupervised phase used a naïve binary label as a placeholder.

Brendan Saltaformaggio, also of Georgia Tech, gave a talk on Retroscope, a system for extracting forensic data from RAM for spatiotemporal data. They interleave execution between a live Android environment with code and data from a memory image to recreate the application's behavior in the past. By reusing the app's own drawing and other internal routines, in conjunction with in-memory data structures that have not been garbage collected, they can re-render screens from the past, even if the application has been closed and logged out of. Because the memory image code knows how to handle the app's data, it can handle all the logistics of rendering the data, and so this method doesn't require deep custom code per application. Brendan demonstrated recovering a deleted draft of a chat from Telegraph after logging out of and closing the app. He's looking at applying this technique to forensics in cases of vehicle or drone-hacking attacks. The talk sparked a lively discussion in the room and on Twitter, as people debated the right way to solve this and the performance implications, such as clearing memory completely on application switch or shut down.

Bayan Bruss from Capital One was next up, talking about financial technology phishing attacks. One out of every 4500 emails is phishing, and email is currently the number one attack vector. Capital One was interested in a solution that would accelerate their analysts and use them more efficiently. They built human-in-the-loop machine learning systems to speed up their analysis and improve defense. They still need MTA filters, which block 98% of attacks, but they couldn't afford to not catch that last 2%. Employees report emails quickly and get rapid feedback from SOC analysts to train both the users and the machine learning. By doing pre-classification, they were able to reduce their analyst workload by 70%. He regards it as empowering your tier 1 analysts by giving them better investigation tools. The goal was not to replace them but to augment them. He emphasized the importance of closing the loop with the analysts and getting the true labels for later retraining. Finally, he talked on the importance of engaging the whole enterprise more effectively. He noted that 64% of phishing drills are recognized, but only 7% of real phishing is, and suggested improving both the quality and frequency of drills. In addition, he noted that it's important to engage users by making it easy to report phishing, giving early feedback and updating the feedback after the analyst looks at it.

Flavio Villanustre from LexisNexus gave a talk on user-entity behavioral analytics (UEBA). His talk was primarily a call to action, covering open problems in UEBA, from dealing with short time series to how to realistically do continuous authentication. He noted that biometric accuracy continues to be quite low, but when used in conjunction with other independent methods, it can strengthen authentication.

Finally, the day closed with a panel session on ML in the world of startups. The panel was composed of Adam Hunt, Chief Data Scientist at RiskIQ; Sven Krasser, Chief Scientist at CrowdStrike; Sean Park, Senior Malware Scientist at Trend Micro; and Kelly Shortridge, Product Manager at SecurityScorecard. Aleatha Parker-Wood moderated and guided the discussion to cover communicating the value of machine learning in a business context, striking a balance between cutting-edge technology and tried-and-true techniques, and what emerging technologies each of them was most excited about. She then offered the closing remarks, thanking the speakers and audience for making ScAINet a success, and encouraging them to form new collaborations and connections within the community.

The consensus from attendees and speakers was that this was a superb lineup of speakers and open discussion, and that they looked forward to larger attendance and more speakers next year.

Special thanks to Google for sponsoring lunch and to the Program and Event committees (Polo Chau, Andrew B. Gardner, Aleatha Parker-Wood, Alina Oprea, Nikolaos Vasiloglou, and Anisha Mazumder) as well as USENIX and organizing staff, including Casey Henderson, Sarah TerHune, and Jenn Hickey.