DEF CON 26 is only one month away! We have a large number of amazing talks planned for everyone in attendance:

Full List

Title	Speaker	Type
The current state of adversarial machine learning	infosecanon	Presentation
StuxNNet: Practical Live Memory Attacks on Machine Learning Systems	Raphael Norwitz	Presentation
Holy BATSense! Deploying TBATS Machine Learning Algorithm to Detect Security Events	Pranshu Bajpai	Presentation
Machine Learning for Network Security Hands-on Workshop: DIYML	Sebastian Garcia	Workshop
Detecting Web Attacks with Recurrent Neural Networks	Fedor Sakharov	Presentation
Using AI to Create Music	dj beep code	Exhibit
It's a Beautiful Day in the Malware Neighborhood	Matt	Presentation
IntelliAV: Building an Effective On-Device Android Malware Detector	Mansour Ahmadi	Presentation
Chatting with your programs to find vulnerabilities	Chris Gardner	Presentation
Hunting the Ethereum Smart Contract: Color-inspired Inspection of Potential Attacks	TonTon Huang	Presentation
Deep Exploit	Isao Takaesu	Exhibit
Beyond Adversarial Learning – Security Risks in AI Implementations	Kang Li	Presentation
DeepPhish: Simulating the Malicious Use of AI	Ivan Torroledo	Presentation
AI DevOps: Behind the Scenes of a Global Anti-Virus Company's Machine Learning Infrastructure	Alex Long	Presentation
Generating Labeled Data From Adversary Simulations With MITRE ATT&CK	Brian Genz	Presentation
JMPgate: Accelerating reverse engineering into hyperspace using AI	Rob Brandon	Presentation
Automated Planning for the Automated Red Team	Andy Applebaum	Presentation
Stop and Step Away from the Data: Rapid Anomaly Detection via Ransom Note File Classification	Mark Mager	Presentation
Identifying and correlating anomalies in Internet-wide scan traffic to newsworthy security events	Andrew Morris	Presentation
Machine Learning Model Hardening For Fun and Profit	Ariel Herbert-Voss	Presentation
The great power of AI: Algorithmic mirrors of society	Aylin Caliskan	Presentation
GAN to the dark side: A case study of attacking machine-learning systems to empower defenses	Li Chen	Presentation
Towards a framework to quantitatively assess AI safety – challenges, open questions and opportunities.	Ram Shankar Siva Kumar	Presentation
Adversarial Patches	Sven Cattell	Presentation
Machine Learning as a Service in Your Pocket	Evan Yang	Exhibit

Panels

(Responsible?) Offensive Machine Learning

Panelists

- @_delta_zero (Moderating)
- @bodaceacat
- @filar
- @Straithe

Abstract:

Cool evil hacks using machine learning are exploding in popularity. Not all ML abuse looks like Terminators, but also DeepFakes, political impersonation, and distracting autonomous cars.

What does “Max Evil” in machine learning look like, really? What ethical boundaries and limitations exist in researching and implementing offensive use cases? Which holes in machine learning systems create the most incidental damage? And what can be done about all this?

Machine Learning and Malware Analysis Panel

Panelists

- @bwall (Moderating)
- @drhyrum
- @gradient_janitor
- @malwareunicorn
- @rharang

Abstract:

Malware classification is one of the most successful uses of Machine Learning in the InfoSec industry. This panel will dive into following topic areas: What is the current state of the art for Machine Learning in Malware Classification? What are the biggest challenges currently present in the field? What does the future of malware classification look like? How can domain experts augment current malware classification efforts? What’s the best way to put humans “into the loop”?

Accepted Talks

The current state of adversarial machine learning

infosecanon

Machine learning is quickly becoming a ubiquitous technology in the computer security space, but how secure is it exactly? This talk covers the research occurring in adversarial machine learning and includes a discussion of machine learning blind spots, adversarial examples and how they are generated, and current blackbox testing techniques.

Heather Lawrence is a cyber data scientist working with NARI. She earned her undergraduate and MS degrees in Computer Engineering from the University of Central Florida focusing on computer security. She is pursuing a PhD in Computer Engineering from the University of Nebraska Lincoln. Her previous experience in cyber threat intelligence modeling, darknet marketplace research, IT/OT testbed development, data mining, and machine learning has led to several awards from capture-the-flag competitions including the National Collegiate Cyber Defense Competition, CSI CyberSEED, and SANS Netwars Tournament. Her current research interests focus on the application of machine learning to cybersecurity problem sets.

StuxNNet: Practical Live Memory Attacks on Machine Learning Systems

Raphael Norwitz

Like all software systems, the execution of machine learning models is dicated by logic represented as data in memory. Unlike traditional software, machine learning systems’ behavior depends less on precise machine opcodes and more on weight vectors and bias parameters. In terms of possible threat models this makes a huge difference: an attacker cannot arbitrarily perturb the program code of a typical executable and have it run error free, however these weight and bias values can be scrambled arbitrarily without any risk of causing a crash. With selective retraining and backpropogation, we demonstrate how one can easily retrain networks to behave completely differently or respond predictably to certain triggers. Thus an attacker looking to compromise an ML system can simply patch these values in live memory, thereby taking control of system with minimal risk of system malfunctions or other detectable side-effects. In this talk, we demonstrate proof of concept malware to attack Neural Networks on Windows 7 and discuss different training paradigms we’ve used. In particular, we focus on minimizing the size of the patch needed to change the model’s behavior, to prove an attack with realistic bounds on network communication. We hope that seeing our malware run against various different frameworks, and showcasing devastating potential of a patched network will provide valuable insight into the mechanisms tomorrow’s hackers could use, and spark a discussion around systems level AI security.

I am a recent graduate from Columbia Univserity with a BA in Computer Science and MS in Machine Learning, and an incoming engineer on the Acropolis Hypervisor team at Nutanix. I have experience with Linux Kernel development, data science and malware analysis. I have interned at Google, Drawbridge and Nimbledroid, and have published research with Columbia’s Wireless and Mobile Networking lab. For fun, I like to be outdoors and train Brazilian Ju-Jitsu.

Holy BATSense! Deploying TBATS Machine Learning Algorithm to Detect Security Events

Pranshu Bajpai

Our “BATSense” security event detection methodology has been running at Michigan State University’s campus for a year and is successfully detecting security anomalies across 300k devices. In this presentation, we will describe the use machine learning, specifically the TBATS forecasting algorithm, to predict future trends for the number of events per second for a variety of device types. The forecasted values are compared against actual observations to alert security personnel of significant deviations. Anomalies are detected based on logs relevant to security events; they may be system modifications, system failures or a football game. Forecasts are never perfect, but when measured over extended use, we have shown that false positives are manageable (1 per week) for true positives of 1 per day. The result a methodology that has been developed and tweaked over time to effectively detect security events, and lessons learned over a year. All arguments presented in this talk will be backed by real world (anonymized) data collected at our university shared with the audience.

Pranshu Bajpai is a security researcher working towards his PhD in Computer Science and Engineering at Michigan State University. His research interests lie in computer and network security, malware analysis, machine learning, privacy, digital forensics, and cyber crimes. In the past, he worked as an independent penetration tester for clients. He has authored several research papers in security magazines and journals and has served as a technical reviewer for books within the security domain. He enjoys working in the security industry and the challenge of testing new technologies for potential weaknesses. In his spare time, he likes solving CTF challenges while listening to classic rock. Connect with him on Twitter: @amirootyet

Detecting Web Attacks with Recurrent Neural Networks

Fedor Sakharov

“Classic Web Application Firewalls (WAFs) mostly use rule-based approach for attack detection. This approach is known to have its pros and cons. Despite offering decent protection from automated attacks and predictable detection results rule-based approach has and always will have certain disadvantages. We all know that it’s useless against 0-day attacks or that even the most sophisticated rules are easily evaded by skilled professionals. That is why a more effective approach should involve some kind of heuristics. Let’s give a chance to artificial intelligence to find something non-obvious for human perception in raw data and try to explain its results.

To this day AI has been more often used for cat classification rather than for detecting application-level attacks on HTTP applications. Our team decided to test the hypothesis that Deep Learning is able to detect web-based attacks effectively. We started with very simple neural network architectures and tried to use them for classification. After some experiments it became clear that we needed more complex networks so we abandoned our attempts to use classification shifting to anomaly detection. Eventually, we ended up using seq2seq model with attention mechanisms which is able to detect zero-day web attacks with minimal number of false positives.”

Irina Stepanyuk is a data scientist from Moscow, Russia. For some time Irina is a researcher in Positive Technologies. She develops data analysis algorithms in relation to information security. Moreover, Irina is a Master’s student in the Faculty of Computer Science at the Higher School of Economics, where she also participates in data science projects and research.

Arseny Reutov is a web application security researcher from Moscow, Russia. Arseny is the Head of Application Security Research at Positive Technologies Ltd where he specializes in penetration testing, the analysis of web applications, and application security research. He is the author of research papers and blog posts on web security published in such magazines as Hacker (Xakep) and HITB Magazine as well as in his blog raz0r.name. He was a speaker at ZeroNights, CONFidence, PHDays and OWASP conferences. Arseny loves making web security challenges (#wafbypass on Twitter) as well as solving them. His passion are modern web technologies and finding vulnerabilities in them.

Fedor is a software developer from Moscow, Russia. He takes interest in various aspects of low-level programming and information security. For some time he has contributed to opensource reverse-engineering framework radare2, his diploma thesis is about transparent application CFG control in runtime and he has a solid experience with Linux kernel programming, drivers as well as kernel subsystems. That’s not all, since recently he leads the security-focused machine learning research at Positive Technologies.”

It’s a Beautiful Day in the Malware Neighborhood

Matt

“Malware similarity analysis compares and identifies samples with shared static or behavioral characteristics. Identification of similar malware samples provides analysts with more context during triage and malware analysis. Most domain approaches to malware similarity have focused on fuzzy hashing, locality sensitivity hashing, and other approximate matching methods that index a malware corpus on structural features and raw bytes. Ssdeep or sdhash are often utilized for similarity comparison despite known weaknesses and limitations. Signatures and IOCs are generated from static and dynamic analysis to capture features and matched against unknown samples. Incident management systems (RTIR, FIR) store contextual features, e.g. environment, device, and user metadata, which are used to catalog specific sample groups observed.

In the data mining and machine learning communities, the nearest neighbor search (NN) task takes an input query represented as a feature vector and returns the k nearest neighbors in an index according to some distance metric. Feature engineering is used to extract, represent, and select the most distinguishing features of malware samples as a feature vector. Similarity between samples is defined as the inverse of a distance metric and used to find the neighborhood of a query vector. Historically, tree-based approaches have worked for splitting dense vectors into partitions but are limited to problems with low dimensionality. Locality sensitivity hashing attempts to map similar vectors into the same hash bucket. More recent advances make the use of k-nearest neighbor graphs that iteratively navigate between neighboring vertexes representing the samples.

The NN methods reviewed in this talk are evaluated using standard performance metrics and several malware datasets. Optimized ssdeep and selected NN methods are implemented in Rogers, an open source malware similarity tool, that allows analysts to process local samples and run queries for comparison of NN methods. “

Matt Maisel is a data scientist passionate about the intersection of machine learning, software engineering, and computer security domains. He’s currently the manager of Security Data Science at Cylance. Matt recently architected a scalable malware analysis and modeling service used to process customer malware detections. He’s worked in several organization within Cylance including research engineering as a software architect and consulting as the technical director of the incident response practice. Matt holds a M.S. in Computer Science with a focus in machine learning and distributed systems from Johns Hopkins University.

IntelliAV: Building an Effective On-Device Android Malware Detector

Mansour Ahmadi

” The importance of employing machine learning for malware detection has become explicit to the security community. Several anti-malware vendors have claimed and advertised the application of machine learning in their products in which the inference phase is performed on servers and high-performance machines, but the feasibility of such approaches on mobile devices with limited computational resources has not yet been assessed by the research community, vendors still being skeptical. In this presentation, we aim to show the practicality of devising a learning-based anti-malware on Android mobile devices, first. Furthermore, we aim to demonstrate the significance of such a tool to cease new and evasive malware that can not easily be caught by signature-based or offline learning-based security tools. To this end, we first propose the extraction of a set of lightweight yet powerful features from Android applications. Then, we embed these features in a vector space to build an effective as well as efficient model. Hence, the model can perform the inference on the device for detecting potentially harmful applications. We show that without resorting to any signatures and relying only on a training phase involving a reasonable set of samples, the proposed system, named IntelliAV, provides more satisfying performances than the popular major anti-malware products. Moreover, we evaluate the robustness of IntelliAV against common obfuscation techniques where most of the anti-malware solutions get affected.”

I am a postdoctoral Research Associate at the Northeastern University. I achieved my Ph.D. from the University of Cagliari. I am co-author of more than 10 research papers mostly about the application of machine learning for malware classification. Two of my works received awards from Kaspersky, and the Anti-Virus I developed received media coverage.

Chatting with your programs to find vulnerabilities

Chris Gardner

During the Cyber Grand Challenge, an automated vulnerability exploitation competition, all the teams used the same approach: use a fuzzer to find bugs, and symbolic execution to generate an exploit for any bugs found. Fuzzers are great at triggering bugs, but their effectiveness is often limited by the quality of the initial testcase corpus that is fed to them. Testcases are easy for humans to create, but hard to generate automatically. Teams used a wide variety of techniques to generate initial seeds: from using very slow symbolic execution techniques to find inputs that triggered execution paths, to just using the word “fuzz” as the seed and hoping for the best. However, many of the programs in the CGC are console programs designed to be used by humans: meaning they give a prompt in English and expect a response. For this research we trained a chatbot Recurrent Neural Network on a set of testcases generated by humans, and ran the RNN against the test set with the goal of finding testcases that had higher code coverage than random guessing and could be used with a fuzzer to find bugs.

Chris recently graduated from UMBC, where he found a passion for malware analysis and binary exploitation. In his spare time he plays CTFs and bikes his way around Washington DC.

Hunting the Ethereum Smart Contract: Color-inspired Inspection of Potential Attacks

TonTon Huang

Blockchain and Cryptocurrencies are gaining unprecedented popularity and understanding. Meanwhile, Ethereum is gaining a significant popularity in the blockchain community, mainly due to the fact that it is designed in a way that enables developers to write decentralized applications (Dapps) and smart contract. This new paradigm of applications opens the door to many possibilities and opportunities. However, the security of Ethereum smart contracts has not received much attention; several Ethereum smart contracts malfunctioning have recently been reported. Unlike many previous works that have applied static and dynamic analyses to find bugs in smart contracts, we do not attempt to define and extract any features; instead we focus on reducing the expert’s labor costs. We first present a new in-depth analysis of potential attacks methodology and then translate the bytecode of solidity into RGB color code. After that, we transform them to a fixed-sized encoded image. Finally, the encoded image is fed to convolutional neural network (CNN) for automatic feature extraction and learning, detecting security flaw of Ethereum smart contract.

Hsien-De Huang (a.k.a. TonTon) is working for Leopard Mobile Inc. (Cheetah Mobile Taiwan Agency), and currently a Ph.D. candidate (IKM Lab.) in the Dept. Computer Science and Information Engineering at National Cheng Kung University, Tainan Taiwan. His research interests include Deep Learning, Blockchain, Malware Analysis, Type-2 Fuzzy Logic, and Ontology Applications, and gave talks at RuxCon 2017, OWASP AppSec USA 2017, Hadoop.TW annual conference 2016, TW CSA Summit 2016 and Hackers in Taiwan Conference (HITCON) 2015 & 2014.

Chia-Mu Yu received his Ph.D degree from National Taiwan University in 2012. He is currently an assistant professor at National Chung Hsing University, Taiwan. He was a research assistant in the Institute of Information Science, Academia Sinica. He was a visiting scholar at Harvard University, Imperial College London, Waseda University, and University of Padova. He was a postdoc researcher at IBM Thomas J. Watson Research Center. He serves as an associate editor of IEEE Access and Security and Communication Networks. His research interests include cloud storage security, IoT security, and differential privacy.

Beyond Adversarial Learning – Security Risks in AI Implementations

Kang Li

A year after we discovered and reported a bunch of CVEs related to deep learning frameworks, many security and AI researchers have started to pay more attention to the software security of AI systems. Unfortunately, many deep learning developers are still unaware of the risks buried in AI software implementations. For example, by inspecting a set of newly developed AI applications, such as image classification and voice recognition, we found that they make strong assumptions about the input format used by training and classifications. Attackers can easily manipulate the classification and recognition without putting any effort in adversarial learning. In fact the potential danger introduced by software bugs and lack of input

validation is much more severe than a weakness in a deep learning model. This talks will show threat examples that produce various attack effects from evading classifications, to data leakage, and even to whole system compromises. We hope by demonstrate such threats and risks, we can draw developers' attention to software implementations and call for community collaborative effort to improve software security of deep learning frameworks and AI applications.

Kang Li is a professor of computer science and the director of the Institute for Cybersecurity and Privacy at the University of Georgia. His research results have been published at academic venues, such as IEEE S&P, ACM CCS and NDSS, as well as industrial conferences, such as BlackHat, SyScan, and ShmooCon. Dr. Kang Li is the founder and mentor of multiple CTF security teams, including SecDawg and Blue-Lotus. He was also a founder and player of the Team Disekt, a finalist team in the 2016 DARPA Cyber Grand Challenge.

DeepPhish: Simulating the Malicious Use of AI

Ivan Torroledo

Machine Learning and Artificial Intelligence have become essential to any effective cyber security and defense strategy against unknown attacks. In the battle against cybercriminals, AI-enhanced detection systems are markedly more accurate than traditional manual classification. Through intelligent algorithms, detection systems have been able to identify patterns and detect phishing URLs with 98.7% accuracy, giving the advantage to defensive teams. However, if AI is being used to prevent attacks, what is stopping cyber criminals from using the same technology to defeat both traditional and AI-based cyber-defense systems? This hypothesis is of urgent importance - there is a startling lack of research on the potential consequences of the weaponization of Machine Learning as a threat actor tool. In this talk, we are going to review how threat actors could exponentially improve their phishing attacks using AI to bypass machine-learning-based phishing detection systems. To test this hypothesis, we designed an experiment in which, by identifying how threat actors deploy their attacks, we took on the role of an attacker in order to test how they may use AI in their own way. In the end, we developed an AI algorithm, called DeepPhish, that learns effective patterns used by threat actors and uses them to generate new, unseen, and effective attacks based on attacker data. Our results show that, by using DeepPhish, two uncovered attackers were able to increase their phishing attacks effectiveness from 0.69% to 20.9%, and 4.91% to 36.28%, respectively.

Ivan Torroledo is the lead data scientist in the Cyxtera Research organization. In this role, he develops and implements Machine and Deep Learning algorithms to enhance phishing detection, network security, fraud detection, and malware mitigation. Ivan is also highly interested in research on the application of Machine and Deep Learning in high energy physics and astrophysics. Before joining Cyxtera, he worked at the Central Bank of Colombia, applying high performance computing tools to monetary policy analysis. He is passionate about applying the most advanced scientific knowledge to cyber security industry. Ivan holds degrees in Economics and Physics.

AI DevOps: Behind the Scenes of a Global Anti-Virus Company's Machine Learning Infrastructure

Alex Long

"Thus far, the security community has treated machine learning as a research problem. The painful oversight here is in thinking that laboratory results would translate easily to the real world, and as such, not devoting sufficient focus to bridging that gap. Researchers enjoy the luxuries of neat bite-sized datasets to experiment upon, but the harsh reality of millions of potentially malicious files streaming in daily soon hits would-be ML-practitioners in the face like a tsunami-sized splash of ice water. And while in research there's no such thing as ""too much"" data, dataset sizes challenge real-world cyber security professionals with tough questions: ""How will we store these files efficiently without hampering our ability to use them for day-to-day operations?"" or ""How do we satisfy competing use-cases such as the need to analyze specific files and the need to run analyses across the entire dataset?"" Or maybe most importantly: ""Will my boss have a heart-attack when he sees my AWS bill?""

In this talk, we will provide a live demonstration of the system we've built using a variety of AWS services including DynamoDB, Kinesis, Lambda, as well as some more cutting edge AWS services such as Redshift and ECS Fargate. We will go into depth about how the system works and how it answers the difficult questions of real world ML such as the ones listed above. This talk will provide a rare look into the guts of a large-scale machine learning production system. As a result, it will give audience members the tools and understanding to confidently tackle such problems themselves and ultimately give them a bedrock of immediately practical knowledge for deploying large-scale on-demand deep learning in the cloud."

Alex Long is currently working as a programmer on the Sophos Datascience Team where he builds tools, scalable backends, and cool visualizations to support the team's research. His latest work has been on creating an online platform for researchers to publish, evaluate, and distribute their latest AI models, thus streamlining the process of productizing AI breakthroughs.

Generating Labeled Data From Adversary Simulations With MITRE ATT&CK

Brian Genz

"Attackers have a seemingly endless arsenal of tools and techniques at their disposal, while defenders must continuously strive to improve detection capabilities across the full spectrum of possible vectors. The MITRE ATT&CK Framework provides a useful collection of attacker tactics and techniques that enables a threat-focused approach to detection.

This technical talk will highlight key lessons learned from an internal adversary simulation at a Fortune 100 company that evolved into a series of data science experiments designed to improve threat detection. "

Brian Genz is a Security Engineer focused on threat hunting, security data science, threat intelligence, and security orchestration, automation & response. He brings experience in the defense intelligence, manufacturing, and financial sectors in the areas of incident response, digital forensics, vulnerability management, and security architecture consulting. He has presented at Derby Con, Circle City Con, CypherCon, the ISSA International Conference, ISACA, InfraGard, and other venues. Brian also serves as adjunct faculty in the information security program at Milwaukee Area Technical College.

JMPgate: Accelerating reverse engineering into hyperspace using AI

Rob Brandon

"One of the most exciting potential applications for artificial intelligence and machine learning is cognitive augmentation of humans. At its best, AI allows humans to handle more information, react faster to complex events, and potentially even sense features of the world that we are currently incapable of perceiving. This has many applications in the security field, such as aiding humans in the task of binary reverse engineering. Reverse engineering binary code is one of the most challenging skill sets in the security field to learn. The ability to look at a block of raw machine code and understand what it does, as well as recognize similarities to code previously seen, often requires years spent doing tedious analysis of large amounts of code.
In this talk I show how we can use machine learning to handle the tedious parts of this process for us. If we show a generative neural network a wide variety of machine code, the network will learn the most relevant features needed to reproduce and describe that code. Once the network is trained, we can show it a new segment of code and capture the state of the neurons at the end of the segment. This neural state is effectively a summary of the entire sequence summarized into a vector. Comparing these vectors allows easy measurement of the similarity of several code sequences by simply measuring the Euclidean distance between them. These vectors can also be used as inputs to other machine learning models that can perform a variety of tasks, such as identifying compiler settings used to generate the code. As part of the presentation, I will also be releasing a tool, the JMPgate framework, which can be used to accomplish tasks like identifying library code within an executable binary. "

Rob is a threat hunter and data scientist with Booz Allen Hamilton's Dark Labs group. He has over 20 years of experience in the tech industry and holds a PhD in computer science from the University of Maryland, Baltimore County. His hobbies include studying the ways that complex systems fall apart and building machines that do his thinking for him so that he can spend more time brewing beer and playing bass.

Automated Planning for the Automated Red Team

Andy Applebaum

"Offensive assessments – i.e., penetration testing, adversary emulation, red teaming – have become a key component of maintaining a secure network. Unfortunately, offensive assessments require significant resources, and can vary in quality and structure based on who specifically is conducting the assessment. In the past few years, we've seen people try to remedy this problem by creating automated offensive assessment tools, but the capabilities and goals of these tools are highly variable, and many either require personnel to manage them or lack the ability to conduct dynamic or end-to-end tests.

We believe that automated offensive assessments can be done better using automated planning. One of the older branches of AI, automated planning seeks to solve problems where an autonomous agent must determine how to compose a sequence of actions together to achieve an objective. Problems in this space can range from constructing offline deterministic plans, to planning under probabilistic conditions, or to planning in scenarios where the world and underlying model are un- or partially-known. Planning techniques have been applied to solve problems in a variety of domains, including controlling unmanned vehicles and designing intelligent agents in computer games.

In this talk, we'll describe how we've leveraged concepts from the automated planning community to help us design CALDERA, a free, open source automated adversary emulation system. Using these concepts, CALDERA dynamically strings techniques – taken from MITRE ATT&CK™ – together to achieve objectives and conduct end-to-end tests. In addition to describing CALDERA itself, we'll also discuss more generally some of the challenges and advantages of deploying automated planning to automated offensive assessments, discussing alternate approaches that we as well as others have considered in tackling this problem. Attendees should walk away with both an understanding of how they can use CALDERA as well as how planning can be used for automated offensive assessments."

Andy Applebaum is a Lead Cyber Security Engineer at MITRE where he works on applied and theoretical security research problems, primarily in the realms of cyber defense, security automation, and automated adversary emulation. Andy has contributed to MITRE's ATT&CK framework and CALDERA adversary emulation platform, as well as other projects within MITRE's internal research and development portfolio. Prior to working at MITRE, Andy received his PhD in computer science from the University of California Davis, where his dissertation topic was using argumentation logic for reasoning in cyber security. Andy's work has been published in multiple conferences and workshops and has most recently spoken at Black Hat Europe. In addition to his PhD, Andy holds a BA in computer science from Grinnell College and the OSCP certification.

Stop and Step Away from the Data: Rapid Anomaly Detection via Ransom Note File Classification

Mark Mager

" The proliferation of ransomware has become a widespread problem culminating in numerous incidents that have affected users worldwide. Current ransomware detection approaches are limited in that they either take too long to determine if a process is truly malicious or tend to miss certain processes due to focusing solely on static analysis of executables. To address these shortcomings, we developed a machine learning model to classify forensic artifacts common to ransomware infections: ransom notes. Leveraging this model, we built a ransomware detection capability that is more efficient and effective than the status quo.

I will highlight the limitations to current ransomware detection technologies and how that instigated our new approach, including our research design, data collection, high value features, and how we performed testing to ensure acceptable detection rates while being resilient to false positives. I will also be conducting a live demonstration with ransomware samples to demonstrate our technology's effectiveness. Additionally, we will be releasing all related source code and our model to the public, which will enable users to generate and test their own models, as we hope to further push innovative research on effective ransomware detection capabilities."

Throughout his career in software engineering and computer security, Mark has served in prominent technical leadership roles in the research and development of advanced computer network operations tools and has provided malware analysis and reverse engineering subject matter expertise to a diverse range of government and commercial clients in the Washington, D.C. metropolitan area.

Identifying and correlating anomalies in Internet-wide scan traffic to newsworthy security events

Andrew Morris

In this presentation, we will discuss using GreyNoise, a geographically and logically distributed system of passive Internet scan traffic collector nodes, to identify statistical anomalies in global opportunistic Internet scan traffic and correlate these anomalies with publicly disclosed vulnerabilities, large-scale DDoS attacks, and other newsworthy events. We will discuss establishing (and identifying any deviations away from) a "standard" baseline of Internet scan traffic. We will discuss successes and failures of different methods employed over the past six months. We will explore open questions and future work on automated anomaly detection of Internet scan traffic. Finally, we will provide raw data and a challenge as an exercise to the attendees.

Andrew Morris is the founder and CEO of GreyNoise Intelligence, a DC-based cyber security company, and likely holds the world record for amount of time staring at Internet-wide scan traffic. Prior to founding GreyNoise, Andrew worked as a researcher, red team operator, and consultant for several large cyber security firms including Endgame, NCC group, and KCG. Outside of work, Andrew enjoys playing fingerstyle acoustic guitar and tries to figure out what his dreams mean.

Machine Learning Model Hardening For Fun and Profit

Ariel Herbert-Voss

Machine learning has been widely and enthusiastically applied to a variety of problems to great success and is increasingly used to develop systems that handle sensitive data - despite having seen that for out-of-the-box applications, determined adversaries can extract the training data set and other sensitive information. Suggested techniques for improving the privacy and security of these systems include differential privacy, homomorphic encryption, and secure multi-party computation. In this talk, we'll take a look at the modern machine learning pipeline and identify the threat models that are solved using these techniques. We'll evaluate the possible costs to accuracy and time complexity and present practical application tips for model hardening. I will also present some red team tools I developed to easily check black box machine learning APIs for vulnerabilities to a variety of mathematical exploits.

Ariel Herbert-Voss is a PhD student at Harvard University, where she specializes in deep learning, cybersecurity, and mathematical optimization. Like many machine learning researchers, she spent plenty of time thinking about deep learning from a computational neuroscience point of view without realizing that skulls make biological neural networks a lot less hackable than artificial ones. Now she thinks about securing deep learning algorithms and offensive applications.

The great power of AI: Algorithmic mirrors of society

Aylin Caliskan

"Following the progress in computing and machine learning algorithms as well as the emergence of big data, artificial intelligence (AI) has become a reality impacting every fabric of our algorithmic society. Despite the explosive growth of machine learning, the common misconception that machines operate on zeros and ones, therefore they should be objective, still holds. But then, why does Google Translate convert these Turkish sentences with gender-neutral pronouns, "O bir doktor. O bir hemşire", to these English sentences, "He is a doctor. She is a nurse"? As data-driven machine learning brings forth a plethora of challenges, I analyze what could go wrong when algorithms make decisions on behalf of individuals and society if they acquire statistical knowledge of language from historical human data.

In this talk, I show how we can repurpose machine learning as a scientific tool to discover facts about artificial and natural intelligence, and assess social constructs. I prove that machines trained on societal linguistic data inevitably inherit the biases of society. To do so, I derive a method that investigates the construct of language models trained on billions of sentences collected from the World Wide Web. I conclude the talk with future directions and open research questions in the field of ethics of machine learning."

Aylin Caliskan is an assistant professor of computer science at George Washington University. Her research interests include the emerging science of bias in machine learning and fairness, AI ethics, data privacy, and security. Her work aims to characterize and quantify aspects of artificial and natural intelligence using a multitude of machine learning and language processing techniques. In her recent publication in Science, she demonstrated how semantics derived from language corpora contain human-like biases. Prior to that, she developed novel privacy attacks to de-anonymize programmers using code stylometry. Her presentations on both de-anonymization and bias in machine learning are the recipients of best talk awards. Her work on semi-automated anonymization of writing style furthermore received the Privacy Enhancing Technologies Symposium Best Paper Award. Her research has received extensive press coverage across the globe, contributing to public awareness on risks of AI. Aylin holds a PhD in Computer Science from Drexel University and a Master of Science in Robotics from the University of Pennsylvania. Before joining the department of computer science at GWU, Aylin was a postdoctoral researcher and a fellow at Princeton University's Center for Information Technology Policy.

GAN to the dark side: A case study of attacking machine-learning systems to empower defenses

Li Chen

"There has been a surge of interest in using machine learning (ML) to automatically detect malware through their dynamic behaviors. These approaches indeed have achieved much higher accurate detection rate and lower false positive rate. ML in threat detection has demonstrated to be a good cop to guard platform security. However should we fully trust ML-powered security? Here, we juxtapose the resiliency and trustworthiness of ML algorithms for security, in the case study of ransomware detection. We propose RD-Fool, an AI-based system to bypass ML-based ransomware detection.

In this talk, we examine the perspectives of ML assuming the role of both a good cop and a bad cop. We first train a variety of deep learning and classical machine learning classifiers for ransomware detection using data collected from file I/O and registry events. We show the classifiers can achieve great performance in terms of classification accuracy and false positive rate for ransomware detection. Then we examine the resiliency of these classifiers using our proposed system RD-Fool. RD-Fool uses random forest and generative adversarial networks (GAN) to generate samples which can bypass the ransomware detectors. We demonstrate both exploratory and causative attacks using RD-Fool, where exploratory attack aims at bypassing the ransomware detector during inference phase, and causative attack aims at poisoning the training data to perturb the ML decision boundary.

The key advantages of RD-Fool include quick identification of the blind spots of the victim ML model and efficient generation of realistic and evasive samples. We examine the quality of the crafted sample using the perturbation distance and the Silhouette score. Our results and discoveries pose interesting and alarming issues such as how much should we trust or utilize ML for better security. “

Li Chen is a data scientist and research scientist in the Security and Privacy Lab at Intel Labs, where she focuses on developing state-of-the-art robust machine learning and deep learning algorithms for security analytics including applications in malware detection and image classification in the adversarial setting. She is also the co-primary investigator (PI) and research lead at the Intel Science & Technology Center for Adversary-Resilient Security Analytics. She designs the roadmaps with Intel and Georgia Tech PIs to jointly meet both industrial and academic research objectives. She provides research direction and in-depth technical guidance to advance the ARSA research agenda. Prior to joining Intel Labs, Li was a Data Scientist in Software and Services Group at Intel, where she focused on developing advanced and principled machine learning methods for cloud workload characterization and cloud computing performance. Li Chen received her Ph.D. degree in Applied Mathematics and Statistics from Johns Hopkins University. Her research interests primarily include machine learning, statistical pattern recognition, random graph inference, data mining, and inference for high-dimensional data. Her research has been featured in a number of pioneering scientific and engineering journals and conferences including IEEE Transactions on Pattern Analysis and Machine Intelligence, Annals of Applied Statistics, Parallel Computing, AAAI Conference on Artificial Intelligence and SPIE. She has given more than 30 technical presentations, including at the Joint Statistical Meeting (the largest statistics conference in North America), AAAI conference, International Joint Conference on Artificial Intelligence, and Spring Research Conference on Statistics and Industry Technology.

Towards a framework to quantitatively assess AI safety – challenges, open questions and opportunities.

Ram Shankar Siva Kumar

While the papers are piling in arxiv on adversarial machine learning, and companies are committed to AI safety, what would a system that assess the safety of ML system look like in practice? Compare a ML system to a bridge under construction. Engineers along with regulatory authorities routinely and comprehensively assess the safety of the structure to attest the bridge’s reliability and ability to function under duress **before** opening it to the public. Can we as security data scientists provide similar guarantees for ML systems? This talk lays the challenges, open questions in creating a framework to quantitatively assess safety of ML systems. The opportunities, when such a framework is put to effect, are plentiful – for a start, we can gain trust with the population at large that ML systems aren’t brittle; that they just come in varying, quantifiable degrees of safety.

Ram Shankar is a Data Cowboy on the Azure Security Data Science team at Microsoft, where his primary focus is modeling massive amounts of security logs to surface malicious activity. His work has appeared in industry conferences like BlueHat, DerbyCon, MIRCon, Strata+Hadoop World Practice of Machine Learning as well as academic conferences like NIPS, IEEE Usenix, ACM - CCS. Ram graduated from Carnegie Mellon University with a Masters in Electrical and Computer Engineering. If you work in the intersection of Machine Learning and Security, he wants to learn about your work!

Adversarial Patches

Sven Cattell

Adversarial examples that fool machine learning systems are a burgeoning field. Academics propose applications in fooling self driving cars or facial recognition systems. However, most of the techniques are useless in the real world. They require precise manipulations to the values of the pixels entering a system. Adversarial patches are an attack that could actually work. One can use them as a sticker on objects that one wants to obscure. This talk will cover how to make them and further applications.

I got my Ph.D. in algebraic topology in 2016 and immediately moved into machine learning to work on something useful to people. I then completed a post-doc in mathematical machine learning where I worked on medical data. I now work at Endgame as a data scientist.

Accepted Exhibits

Using AI to Create Music

dj beep code

Creativity has traditional been a purely human pursuit. However, with recent developments in computational creativity, it has become apparent that the generation of art can now be accelerated with artificial intelligence. Come prepared to learn about reinforcement learning, deep belief networks, and be entertained by music composed in mere seconds, right in front of your eyes.

“With a BS in Applied Mathematics, and a MS in computer science, dj launched her career in engineering in working on the Thirty Meter Telescope project. Over the next 12 years, she specialized in remote sensing algorithms, culminating as the principle investigator in an Office of Naval Research contract on the classification of signals. In 2014 she took her breadth of machine learning knowledge in applied research to the IBM Watson group. Within IBM Watson she leads teams that create AI applications for business, and on the side hacks on The Watson Beat code base. She also plays bass guitar in a bad cover band.”

Deep Exploit

Isao Takaesu

DeepExploit is fully automated penetration tool linked with Metasploit. It identifies the status of all opened ports on the target server and executes the exploit at pinpoint using Machine Learning.

Isao Takaesu is CISSP. He is working in Mitsui Bussan Secure Directions, Inc. as security engineer and researcher. He found many vulnerabilities in client's server and proposed countermeasures to client. He thinks that there's more and wants to find vulnerabilities. Therefore, he is focused on artificial intelligence technology for cyber security. Now, he is developing the penetration test tool using machine learning.

Machine Learning as a Service in Your Pocket

Evan Yang

"If you struggle with building a machine learning (ML) classifier for the data, this Machine Learning as a Service (MLaaS) is a quick and handy solution for you. Originally designed for security researcher, now this feature packed service was open sourced to public. This service can take time-series data, such as API log etc., to generate ML models with few mouse clicks. The graphic user interface could guide you through the ML pipeline steps, visualize the performance and help to optimize the ML model. The unique feature analysis tool allow to drill down individual samples and to tune the ML model in a security perspective way."

Evan Yang is a security researcher in Intel Privacy & Security Lab. He had worked on Windows and Android security related topics for past few years. His latest focus is around the deep learning application on Windows ransomware. He also had been a database architect and software developer to provide solutions and build applications in production.

Accepted Workshops

Machine Learning for Network Security Hands-on Workshop: DIYML

Sebastian Garcia

Creating new Machine Learning algorithms with the new frameworks its easier than ever. However, our models still need designing, evaluation, tuning and specially good datasets. In this workshop we will share high-quality and real datasets of normal users working in their computers while being attacked and infected with malware. The goal is to learn to understand the problem, label data, identify features, create your own ML model and finally test it against all the other models in the room! A fast-paced workshop going from traffic understanding to working python ML models in 2hs. Learn why ML is so difficult and so useful. Work in teams to obtain the highest detection performance and improve your knowledge.
Python/NetFlows/Bro/SciKit/pandas/TensorFlow, use what you need!

Sebastian is a malware researcher and security teacher that has extensive experience in machine learning applied on network traffic. He created the Stratosphere IPS project, the first machine learning-based, free-software IPS. Its goal its to protect the civil society. As a researcher in the Artificial Intelligence group of the Czech Technical University in Prague, he believes that free software and machine learning tools can help better protect users from the abuse of their digital rights. He has been teaching in several countries and Universities and working on penetration testing for both corporations and governments. He was lucky enough to talk and give workshops in CCC, BSides Budapest, Ekoparty, DeepSec, Hacktivity, Botconf, Hacklu, InBot, SecuritySessions, ECAI, CitizenLab, ArgenCor, Free Software Foundation Europe, VirusBulletin, BSides Vienna, HITB Singapore, CACIC, etc. As a co-founder of the MatesLab hackspace he worked on honeypots, malware detection, distributed scanning (creator of dnmap) keystroke dynamics, Bluetooth analysis, privacy protection, intruder detection, robotics, microphone detection with SDR (Salamandra) and biohacking. He is also a proud co-founder of the Independent Fund for Women in Tech.

Closing Notes

Our Call for Volunteers is open! If you want to help our village be a success, sign up here