Extracting Machine Learning Features to Detect Malicious HTTPS Traffic

(Technical Report)

Maintaining Accountability in a Criminal Justice System that Uses Machine Learning

(STS Research Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Adam Klein
Spring, 2020

Technical Project Team Members
NA

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Introduction

With the advent of machine learning, society has gained the ability to extract previously unseen insights from data. Many believe that these insights can enhance our ability to make pressing decisions in a wide range of areas. One example of this is in the identification of cyber attacks. In 2017, an attack that used ransomware called WannaCry locked access to computing devices throughout the world until ransom payments were made. The attacks targeted key institutions, such as hospitals and banks, demonstrating how modern cyber attacks can create a crippling impact on society (Perlroth & Sanger, 2017). Given these stakes, experts have proposed that machine learning can be used to enhance our ability to identify such threats in advance (Oprea, Li, Norris, & Bowers, 2018). A specific challenge that this could be useful for is identifying encrypted malware communications over the HTTPS protocol (Strasak, 2017). My technical research will investigate whether network logs can produce a set of machine learning features that help prevent attacks by identifying malicious HTTPS traffic

A similar threat-detection strategy is being used in the realm of criminal justice. Algorithms are being used to predict crime and forecast recidivism (Liptak, 2017). However, studies have revealed that these techniques can promote racially discriminative practices (Lum & Isaac, 2016). Furthermore, there is a question of whether a black-boxed algorithm erodes accountability by making the criminal justice system less transparent (Ananny & Crawford, 2018). Becoming reliant on this technology could change how a society decides who goes to jail and how it holds the people who decide this accountable. My STS research will investigate the problem of how accountability can be maintained in a criminal justice system that makes decisions informed by machine learning algorithms. Together, these studies will examine how machine learning can be leveraged to gain actionable insights, but also where it can inhibit the

ability of a society to maintain its social values.

Technical Topic

My technical thesis investigates the problem of finding machine learning features that identify malware HTTPS traffic. HTTPS is a protocol for secure network communication. Since HTTPS is encrypted, malware has started to use it to communicate across networks without being detected (Strasak, 2017). Large enterprises are already having difficulty identifying threats as their networks become more complex (Oprea et al., 2018). Malware communication over HTTPS makes attack prevention more difficult (Strasak, 2017). The potential impacts of modern cyber attacks were witnessed in 2017, when ransomware called WannaCry locked access to computing devices in hospitals in England (Bilefsky, 2017). Since such attacks put human lives at risk, it is essential to develop tools that can detect them in advance.

One approach to detecting malware HTTPS traffic is documented in a thesis from Czech Technical University. This method uses data from network traffic logs to identify encrypted malware traffic (Strasak, 2017). This research is less reliable than other sources since it is an undergraduate thesis. However, it is useful since it details a process that we can use as a starting point in our research. Rather than taking its results at face value, we can recreate it and investigate how well this existing model works. Furthermore, Strasak's methodology is similar to the methodology employed in a number of peer-reviewed papers. One paper presents a system called Beehive that generates machine learning features from network log data (Yen et al., 2013). Another paper proposes a system called MADE that applies machine-learning techniques to an enterprise's network logs to detect potentially malicious activity (Oprea et al., 2018). While these papers do not specifically address the issue of malware communications over HTTPS, they highlight how machine learning can be used to analyze network logs for anomalous behaviors.

I agree with the consensus in these papers that analyzing network logs is a promising approach to detecting threats on computer networks. The papers also provide a proof-of-concept for how this can be applied to enterprise networks that produce immense quantities of data. This provides insight into how to build a system for identifying malware HTTPS traffic in a setting with many users making many HTTPS requests, such as the University of Virginia. To this end, my project will emulate these approaches by using logs of normal and malware HTTPS traffic to identify features that distinguish the two. These features can then be fed to a machine-learning algorithm that makes a decision about what is normal and what is malware. I am working with Professor Malathi Veeraraghavan in the Department of Electrical Engineering to produce a set of machine-learning features that can accurately inform this decision.

STS Topic

My STS thesis investigates whether the use of algorithmic policing technology makes actors in the United States' criminal justice system less accountable by providing them with an opaque mathematical process to rationalize their actions. The criminal justice system in the United States produces racially disproportionate outcomes, with black and Hispanic populations being incarcerated at notably higher rates than white populations (Carson, 2018). The data sets utilized by predictive policing algorithms are biased by these outcomes, and therefore have the potential to exacerbate this systemic discrimination (Kirkpatrick, 2017). A society needs to be able to hold its criminal justice systems accountable to prevent discriminative practices. It is therefore necessary to study whether a system can be held accountable when it makes decisions inside an algorithmic black box.

The majority of the current research on this subject does not address accountability, but rather examines whether predictive policing software and algorithmic recidivism forecasters

promote racial discrimination in the criminal justice system. Lum and Isaac use statistical simulations to demonstrate how PredPol, a leading predictive policing software package, would enhance existing biased police practices (Lum & Isaac, 2016). Dressel and Farid even go as far as to call the efficacy of this technology into question, conducting an experiment to show that COMPAS, a recidivism forecaster, generated predictions that were no more accurate than those produced by people with no knowledge of the criminal justice system (Dressel & Farid, 2018). These papers are peer-reviewed and authored by experts in statistics and computer science, lending credibility to the concerns about this technology. I agree that using predictive policing software can be discriminative if it is informed by biased data. However, I would argue that there is a more important question of whether it is even possible to audit for such practices amongst law enforcement agencies that use this technology. The aforementioned studies point out ways in which these algorithms can fail. If we start to blindly trust the decisions made by these algorithms, we may come to view unjust law enforcement practices as scientifically justified. Moreover, if actors in the criminal justice system can outsource part of their decisions to a black box, the standard by which we expect these actors to justify their actions may be lowered.

There is not extensive research into how the software in question may impact the accountability of law enforcement and judicial agencies. However, Ananny and Crawford have discussed the concept of transparency as it relates to accountability. Ananny and Crawford argue that being able to observe the way a system works does not necessarily yield an understanding of that system, particularly in systems with algorithmic components (Ananny & Crawford, 2018). As evidence, they point out that engineers sometimes do not know how their own codes works, especially in the field of machine learning (Ananny & Crawford, 2018). This implies that understanding algorithms that aid in decision processes, and therefore holding the resulting

decisions accountable, is a potentially intractable problem. In the realm of algorithms, transparency alone fails to provide accountability. This begs the question of whether there exists a framework that would allow us to regulate algorithmically informed decisions.

I am going to organize my research on this topic through Actor-Network Theory, which helps characterize the relationships between technology, people, and ideas. By observing the social structures that arise between actors, we can analyze how a technology could impact stakeholders. The criminal justice system is already a complicated web of humans, organizations, and technologies. The use of predictive software introduces new actors into this system. The algorithms and the companies that develop them are two examples that are immediately apparent. However, some authors have argued that the primary concerns with this technology arise from data sets that are biased by existing law enforcement practices (Kirkpatrick, 2017). Thus, the data sets and the groups that collect them are also important actors. By analyzing the current state of this actor-network, it is possible to determine whether the introduction of these new actors creates any fundamental conflicts with the current standard for accountability in the criminal justice system.

I plan to research this topic by looking at areas where predictive policing conflicts with legal precedents and historical notions of accountability. For example, the Supreme Court has ruled that being in a "high-crime area" is relevant context to inform a police officer's "reasonable suspicion" to search someone (*Illinois v. Wardlow*, 1999). Opponents of predictive policing accuse it of creating "feedback loops" that are biased towards certain communities, since greater police presence leads to more arrests and the designation of a "high-crime area" (Bennett Moses & Chan, 2018). If predictive policing technology informs the basis for a search, how does society validate the soundness of this decision? As this case demonstrates, human

accountability is integral to our understanding of certain rights. By studying this history, we can establish the algorithmic black box as an actor in the criminal justice system and then evaluate whether it fundamentally conflicts with the legal precedents used in the United States to justify sending someone to prison.

Conclusion

At the end of my technical project, I plan to deliver a set of machine learning features that effectively detect malicious HTTPS traffic. On the completion of my STS project, I intend to deliver an enhanced understanding of whether predictive policing technology can be utilized without eroding the accountability of the criminal justice system. These two use cases for machine learning are closely tied together in that they both aim to predict threats to society in advance. However, the issue of accountability in the criminal justice system demonstrates that this approach may not be equally applicable to all problems. Together, these projects will help define a boundary between where machine learning offers new insight and where it obscures important social decisions.

References

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Bennett Moses, L., & Chan, J. (2018). Algorithmic prediction in policing: assumptions, evaluation, and accountability. *Policing and Society*, *28*(7), 806–822. https://doi.org/10.1080/10439463.2016.1253695

Berk, R. (2017). An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. *Journal of Experimental Criminology*, *13*(2), 193–216. https://doi.org/10.1007/s11292-017-9286-2

Bilefsky, D. (2017, December 22). British Patients Reel as Hospitals Race to Revive Computer Systems. *The New York Times*. Retrieved from https://www.nytimes.com/2017/05/13/world/europe/uk-hospitals- cyberattack.html

Carson, E. A. (2018). *Prisoners in 2016* (No. NCJ 251149). Retrieved from Bureau of Justice Statistics website: https://www.bjs.gov/index.cfm?ty=pbdetail&iid=6187

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, *4*(1), eaao5580. https://doi.org/10.1126/sciadv.aao5580

Illinois v. Wardlow. , 528 US 119 (Supreme Court 1999).

Kirkpatrick, K. (2017). It's not the algorithm, it's the data. *Communications of the ACM*, *60*(2), 21–23. https://doi.org/10.1145/3022181

Liptak, A. (2017, December 22). Sent to Prison by a Software Program's Secret Algorithms. *The New York Times*. Retrieved from https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software- programs-secret-algorithms.html

Lum, K., & Isaac, W. (2016). To predict and serve? *Significance*, *13*(5), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized Controlled Field Trials of Predictive Policing. *Journal of the American Statistical Association*, *110*(512), 1399–1411. https://doi.org/10.1080/01621459.2015.1077710

Oprea, A., Li, Z., Norris, R., & Bowers, K. (2018). MADE: Security Analytics for Enterprise Threat Detection. *Proceedings of the 34th Annual Computer Security Applications Conference on - ACSAC '18*, 124–136. https://doi.org/10.1145/3274694.3274710

Perlroth, N., & Sanger, D. E. (2017, December 22). Hackers Hit Dozens of Countries Exploiting Stolen N.S.A. Tool. *The New York Times*. Retrieved from https://www.nytimes.com/2017/05/12/world/europe/uk-national-health- service-cyberattack.html

Strasak, F. (2017). *Detection of HTTPS Malware Traffic*. Czech Technical University in Prague.

Yen, T.-F., Oprea, A., Onarlioglu, K., Leetham, T., Robertson, W., Juels, A., & Kirda, E. (2013). Beehive: large-scale log analysis for detecting suspicious activity in enterprise networks. *Proceedings of the 29th Annual Computer Security Applications Conference on - ACSAC '13*, 199–208. https://doi.org/10.1145/2523649.2523670