Elizabethtown College JayScholar

Landmark Conference Summer Research Symposium Landmark Conference Summer Research Symposium 2019

Jul 11th, 1:30 PM - 2:30 PM

Applying Machine Learning to Encrypted Network Traffic for Malware Detection

Derek Manning *Elizabethtown College,* manningd1@etown.edu

Follow this and additional works at: https://jayscholar.etown.edu/landmark Part of the <u>Computer Sciences Commons</u>

Recommended Citation

Manning, Derek, "Applying Machine Learning to Encrypted Network Traffic for Malware Detection" (2019). *Landmark Conference Summer Research Symposium*. 12. https://jayscholar.etown.edu/landmark/2019/july11/12

This Event is brought to you for free and open access by the Programs and Events at JayScholar. It has been accepted for inclusion in Landmark Conference Summer Research Symposium by an authorized administrator of JayScholar. For more information, please contact kralls@etown.edu.

Applying Machine Learning to Encrypted Network Traffic for Malware Detection

Derek Manning

Mentor: Dr. Peilong Li



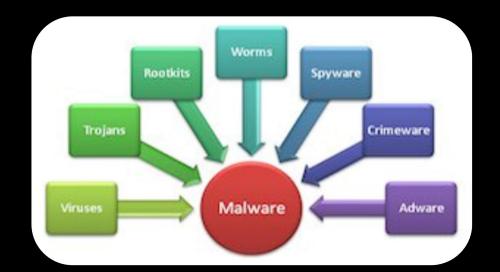
Background

- Malware: malicious software with intent to cause damage
- Large expense for businesses
- \$55 Billion in damages every year
- 130 large-scale, targeted breaches in the U.S. annually
 - Growing 27 percent per year



Intro to Malware Analysis

- Take advantage of vulnerabilities in a system
- Offline analysis and inference
- Rule based security
- Deep packet inspection



Motivation

- Real-time inference with high accuracy
- Optimize for Intel hardware
- On encrypted traffic

Over 50% of the traffic on the internet is encrypted



Source: Sandvine Global Internet Phenomena Report, Oct, 2018

Tools for Analysis

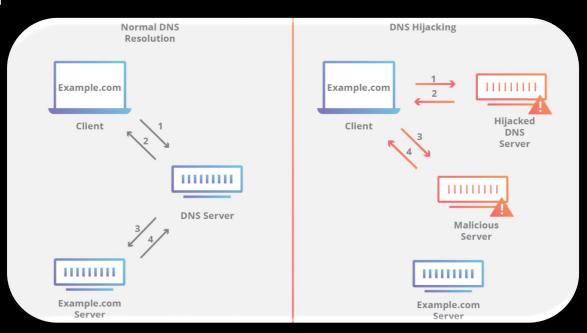
- Packet capture tool
- Cisco Joy (packet cleaning)
- Python
- Intel Data Analytics Acceleration Library (DAAL)





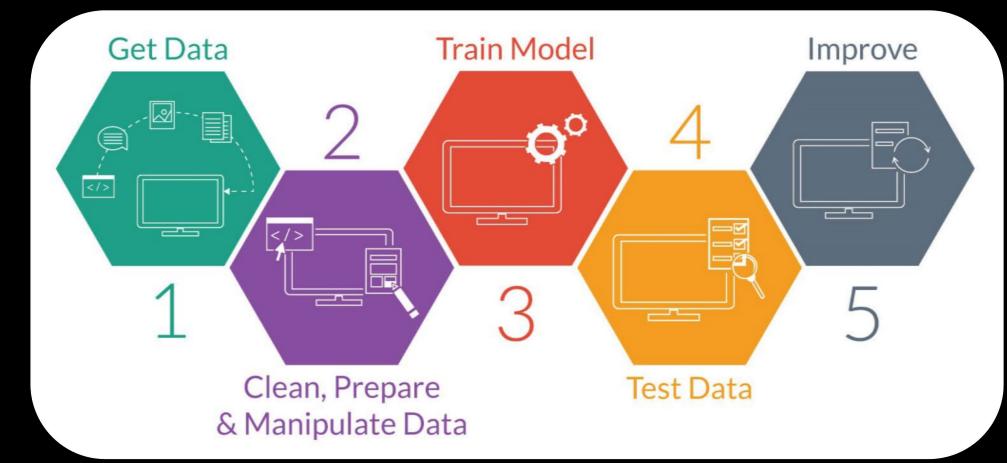
Network Traffic Structure

- HTTP (Hypertext Transfer Protocol)
- TLS (Transport Layer Security)
- DNS (Domain Name System)
- Encrypted data



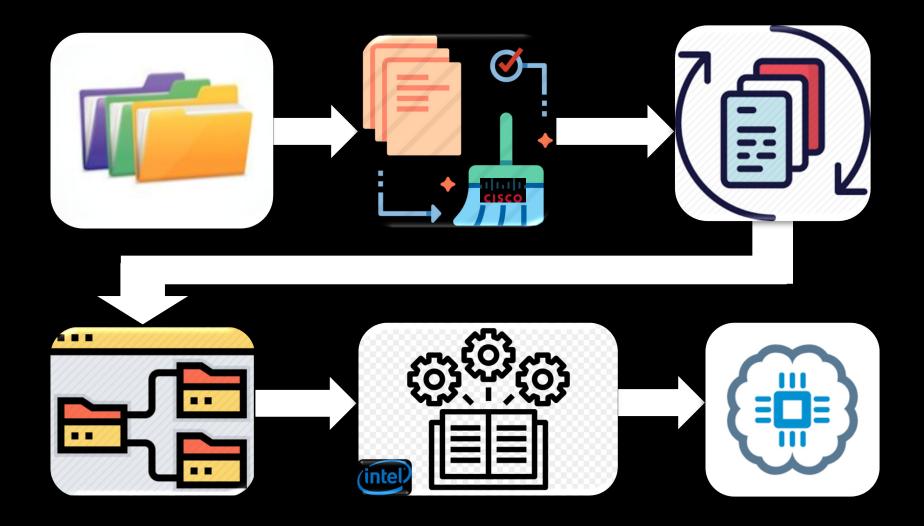
Machine Learning for Inference

Method of data analysis that automates analytical model building



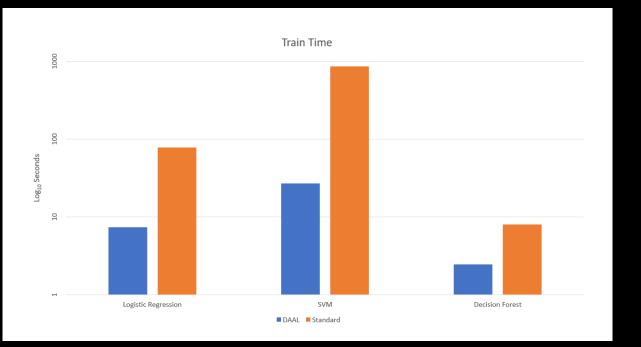


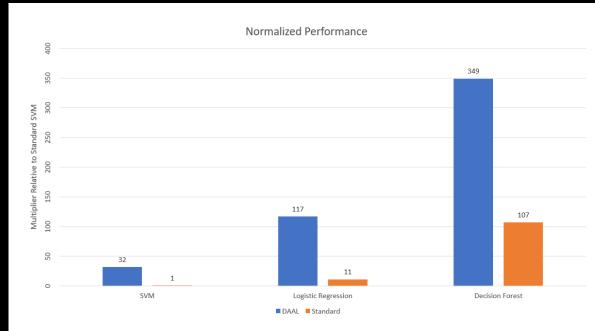
Overall Design



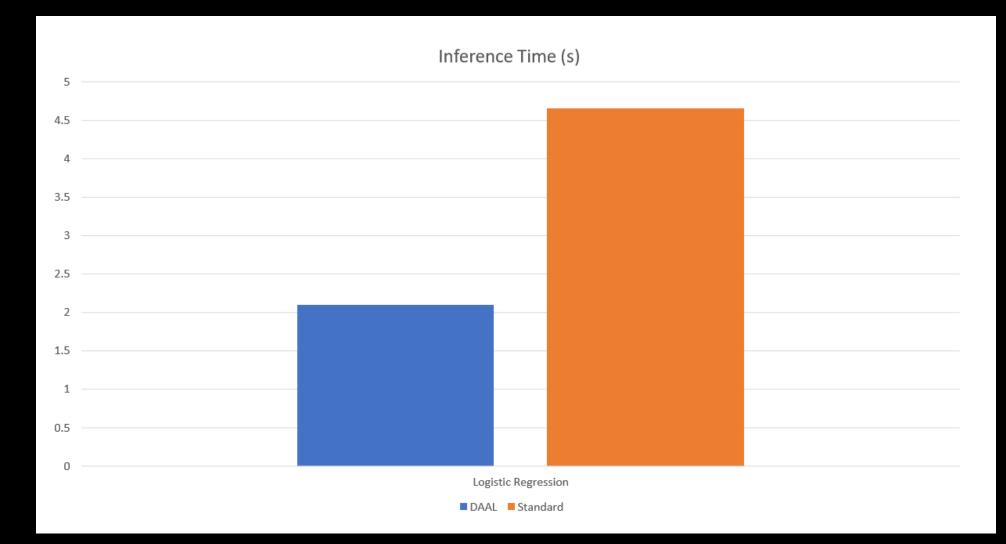
Intel DAAL Background and Results

- Optimizes code for underlying hardware
- Key Takeaway: DAAL 10X faster on average





Inference



Summary

- Prevalence of malware
- Packet metadata as features, not actual data
- Form a dataset, train model
- Optimize models using Intel DAAL
- Apply model to new data

Q & A

References

- <u>https://www.webfx.com/blog/internet/cost-of-computer-viruses-infographic/</u>
- <u>https://www.varonis.com/blog/cybersecurity-statistics/</u>
- <u>https://www.forbes.com/sites/forbestechcouncil/2018/09/28/breaking-down-malware-why-its-still-one-of-the-biggest-threats-facing-businesses/#40765e1afe1a</u>
- <u>https://www.datex.ca/blog/9-types-of-malware-and-how-to-recognize-them</u>
- <u>https://www.novainfosec.com/2013/12/23/malware-analysis-and-incident-response-for-the-lazy/</u>
- <u>https://www.sandvine.com/blog/global-internet-phenomena-encrypted-traffic-dominates-the-internet</u>
- <u>https://www.cloudflare.com/learning/dns/dns-security/</u>
- <u>https://www.sas.com/en_us/insights/analytics/machine-learning.html</u>
- <u>https://blog.usejournal.com/machine-learning-for-beginners-from-zero-level-8be5b89bf77c</u>
- <u>https://www.python.org/community/logos/</u>
- <u>https://commons.wikimedia.org/wiki/File:Intel-logo.svg</u>

Additional Data

