Summer 2019

Applying Machine Learning to Encrypted Network Traffic for Malware Detection

Derek Manning

Applying Machine Learning to Encrypted Network Traffic for Malware Detection

Derek Manning

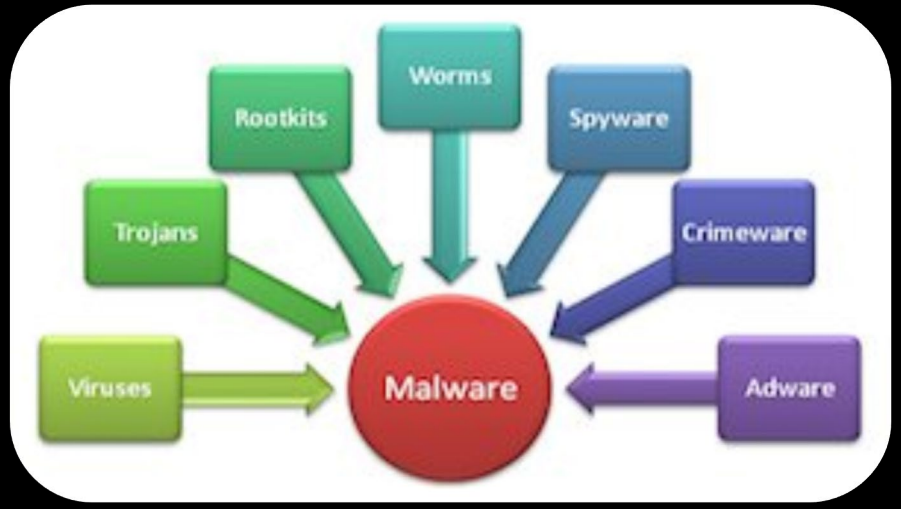
Mentor: Dr. Peilong Li

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

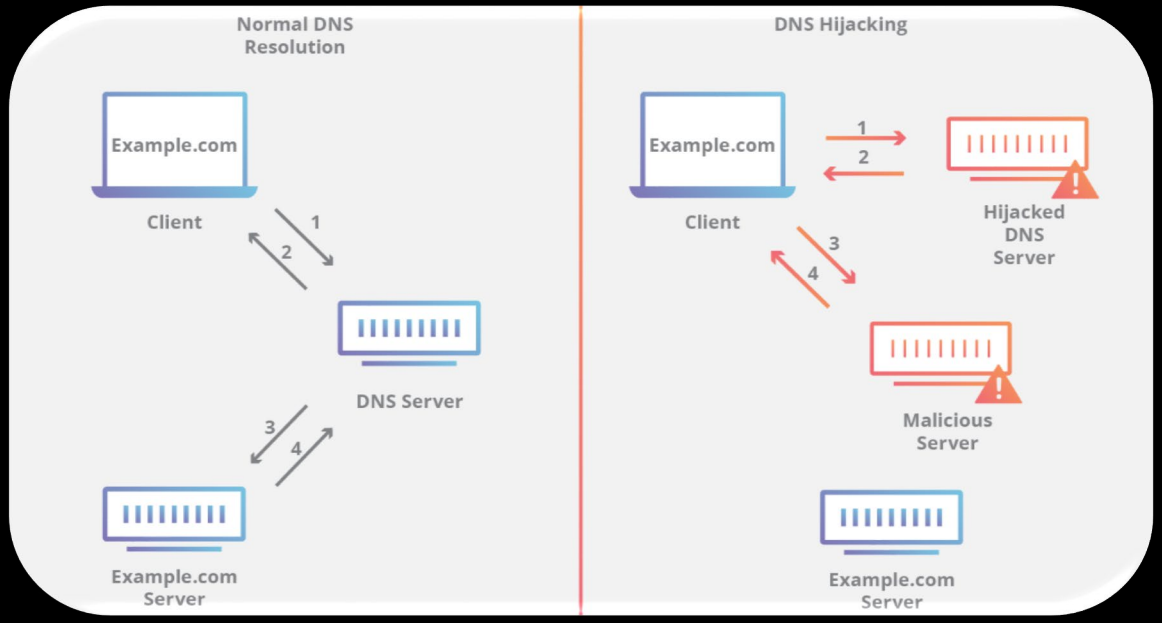
- Real-time inference with high accuracy
- Optimize for Intel hardware
- On encrypted traffic

Tools for Analysis

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

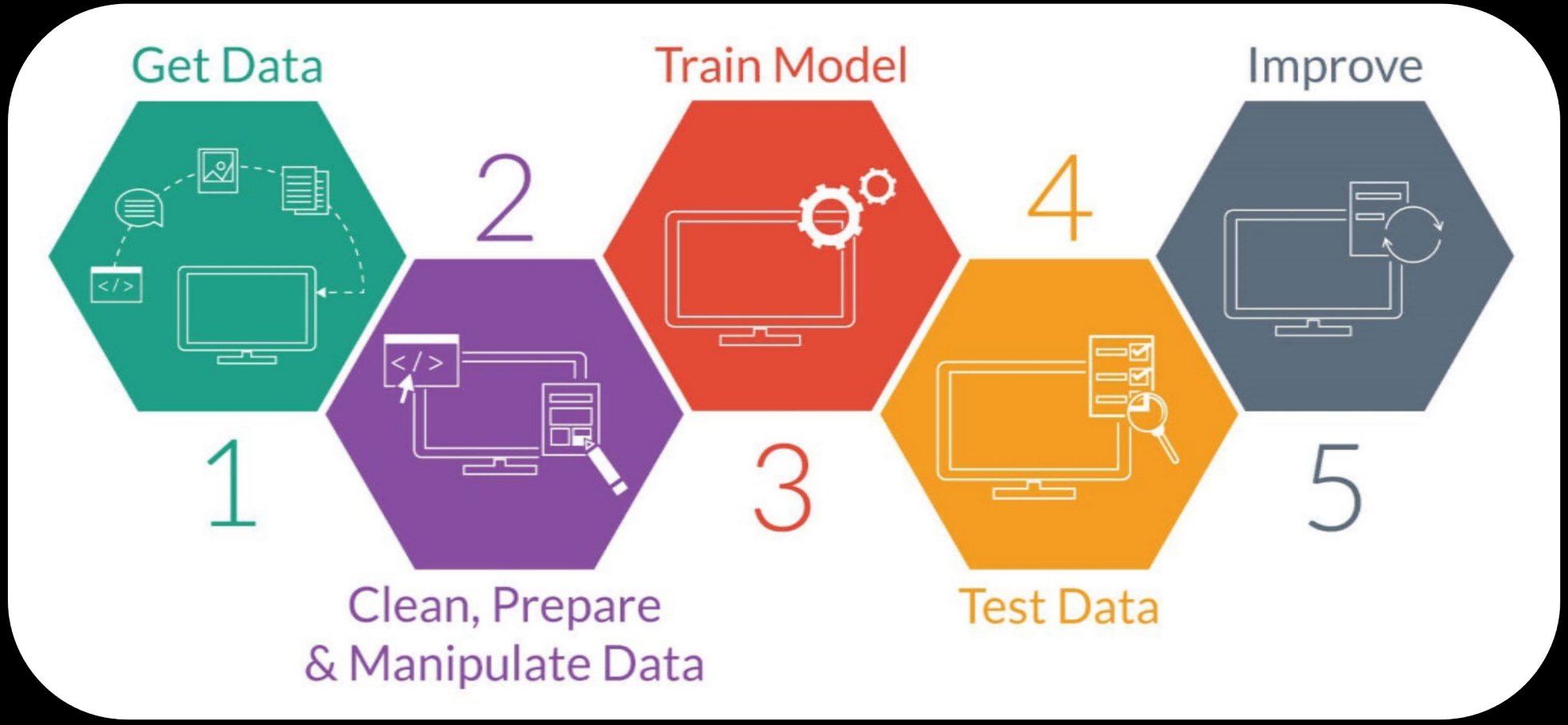
Network Traffic Structure

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

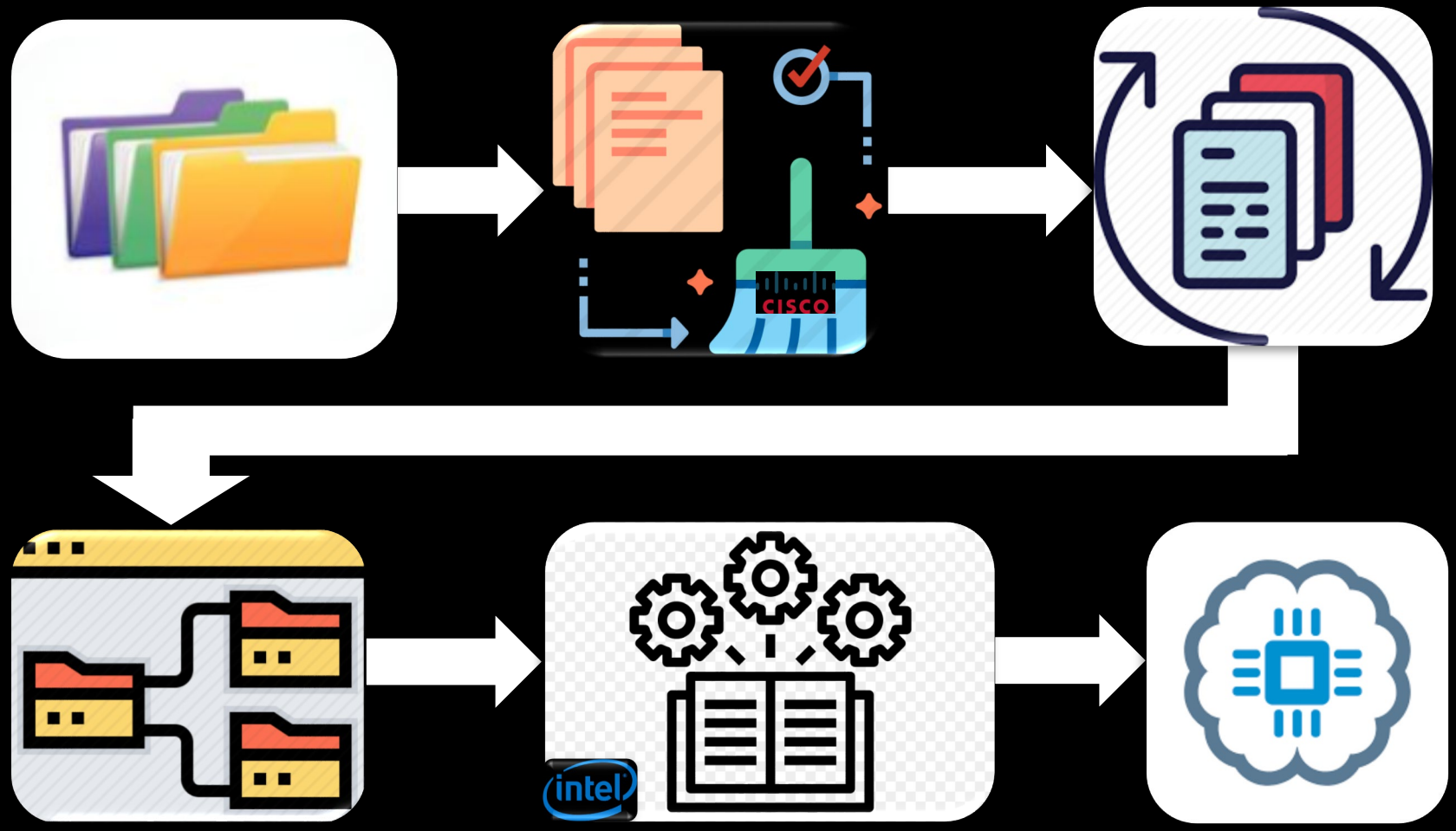


Machine Learning for Inference

- Method of data analysis that automates analytical model building

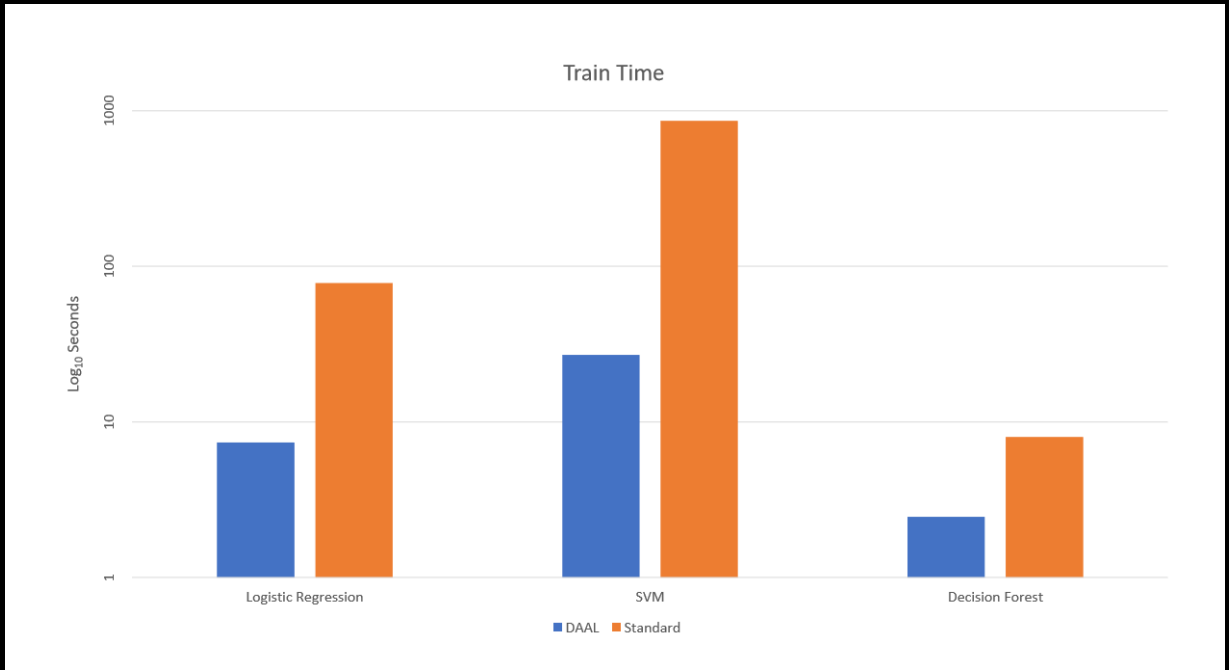


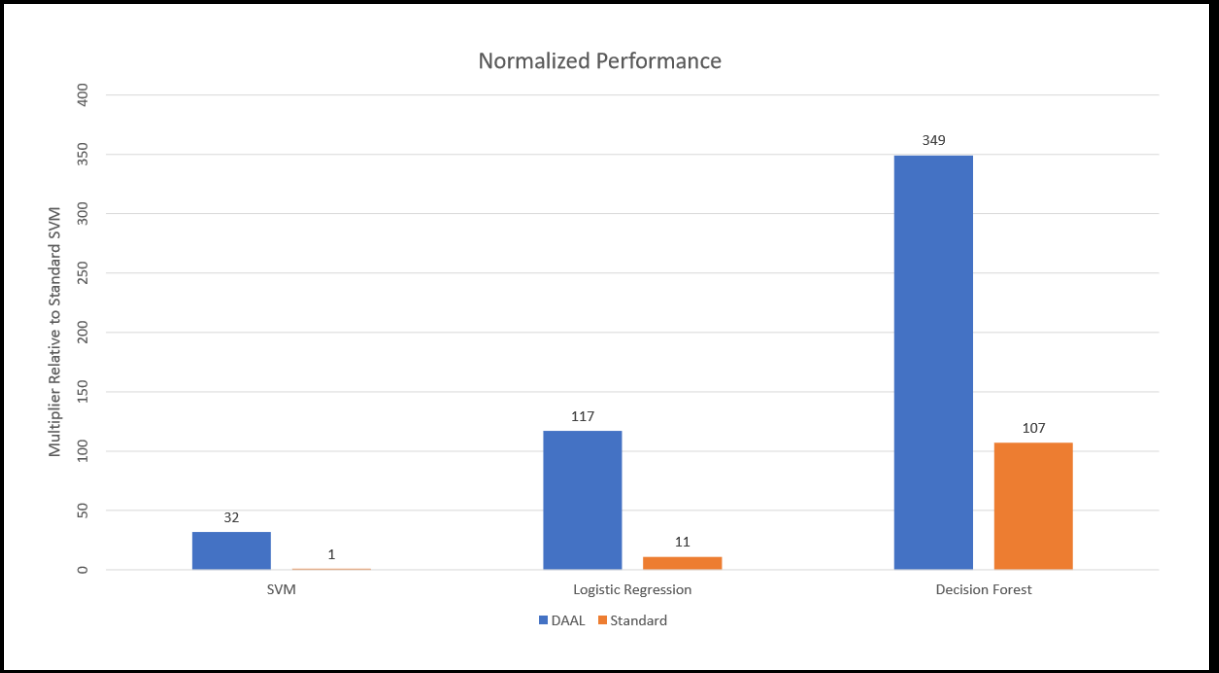
Overall Design



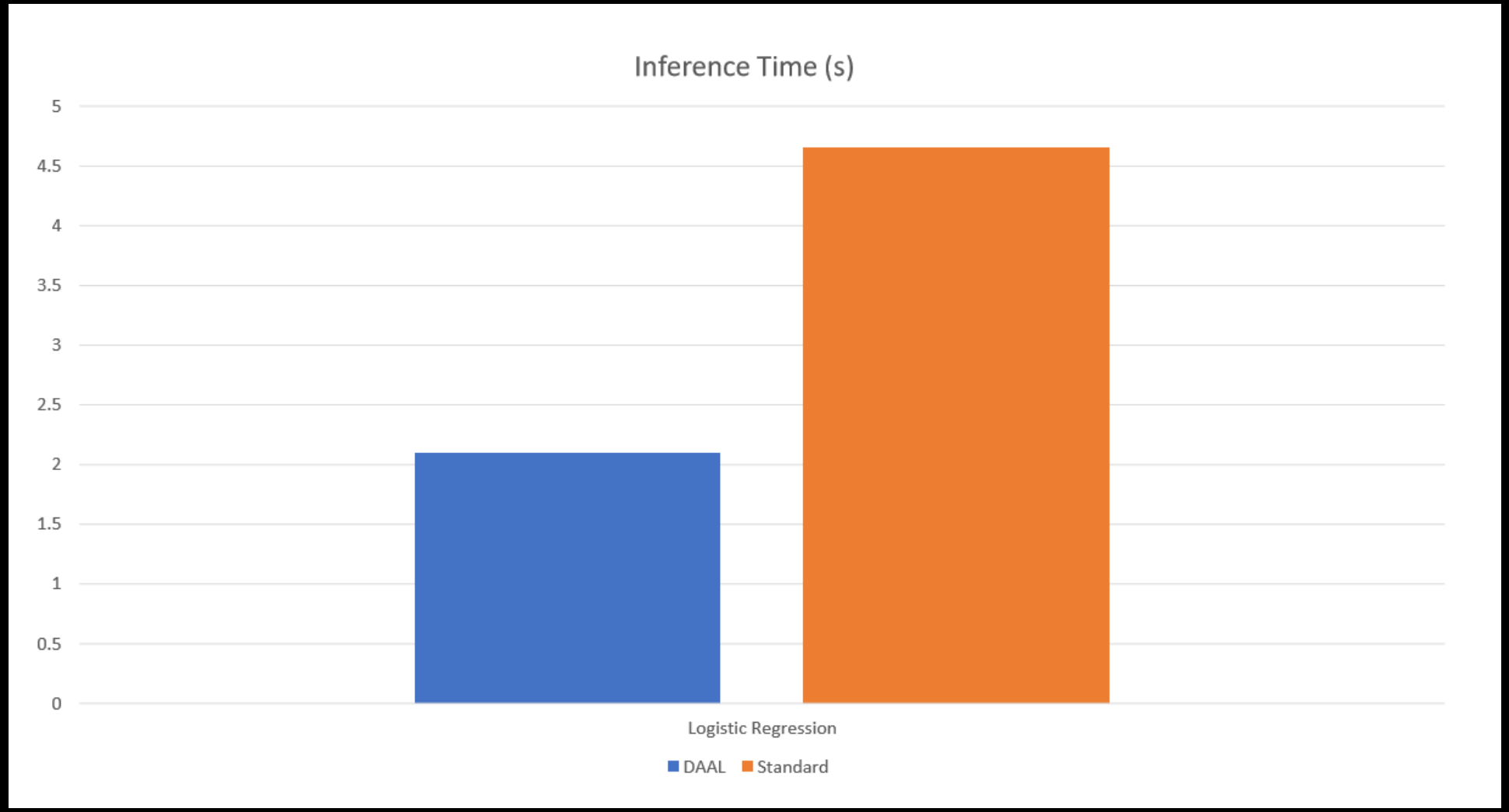
Intel DAAL Background and Results

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





Inference


Inference Time (s)
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Logistic Regression
DAAL
Standard

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

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

Additional Data