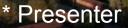


Machine Learning-Based Detection of Ransomware Using SDN

Greg Cusack*, Oliver Michel, Eric Keller 3/21/18



Overview

- Ransomware Overview
- Previous work
- Programmable Forwarding Engines (PFEs)
- Method
- Classification
- Results
- Current Progress



Ransomware

- Malicious software holds a victim's files at ransom
- Files held until ransom paid
- Two main types:
 - Locker
 - Crypto
- Difficult to develop long term solutions
- IoT boom -> More avenues for infection
- Ransomware as a Service (RaaS)

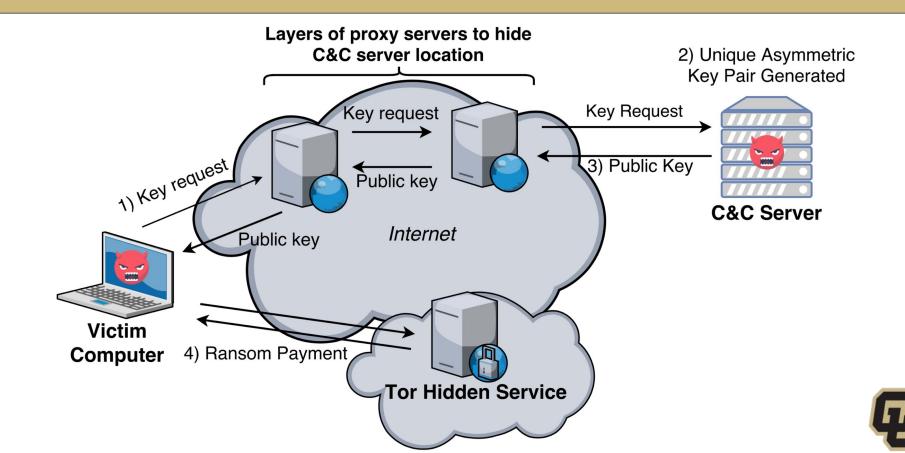
WannaCry Ransomware

Wanna Decryptor 1.0	<u>×</u>
	Ooops, your files have been encrypted!
	What Happened to My Computer?
	Your important files are encrypted.
	Many of your documents, photos, videos, databases and other files are no longer accessible because they have been encrypted. Maybe you are busy looking for a
Payment will be raised on	way to recover your files, but do not waste your time. Nobody can recover your files without our decryption service.
5/15/2017 16:25:02	
Time Left	Can I Recover My Files?
02:23:58:28	Sure. We guarantee that you can recover all your files safely and easily. (But you have not so enough time.)
	You can try to decrypt some of your files for free. Try now by clicking <decrypt>.</decrypt>
Your files will be lost on	If you want to decrypt all your files, you need to pay.
5/19/2017 16:25:02	You only have 3 days to submit the payment. After that the price will be doubled. Also, if you don't pay in 7 days, you won't be able to recover your files forever.
Time Left	Also, il you don't pay in r days, you won't be able to recover your lifes lorever.
06:23:58:28 📕	How Do I Pay?
	Send \$300 worth of bitcoin to this address:
About bitcoin	Bitcoin ACCEPTED HERE 152GqZCTcys6eCjDkE3DypCjXi6QWRV6V1
How to buy bitcoins?	
Contact Us	Check <u>P</u> ayment <u>D</u> ecrypt

Source: https://www.pbs.org/newshour/science/everything-need-know-wannacrypt-ransomware-attack

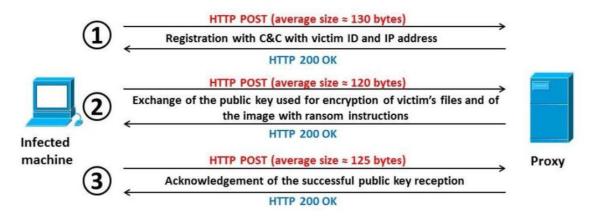


Ransomware Data Flow



Previous Work

- EldeRan: Machine learning approach for ransomware classification
 - Track Windows API calls, file system operations, registry key operations, etc.
- Software-defined networking-based detection of crypto ransomware
 - Fingerprint HTTP traffic
- Most packet trace approaches are payload-based





Source: K. Cabaj, M. Gregorczyk, and W. Mazurczyk. Software-defined networking-based crypto ransomware detection using http traffic characteristics. arXiv preprint arXiv:1611.08294, 2016

Programmable Forwarding Engines (PFEs)

- High-rate, programmable, network switches
- Supports the scalable generation of rich flow records
- Can process network data at high-rates of speed and extract vital, per-packet flow information
- Provides data and speed necessary for network, flow-based, traffic analysis and fingerprinting



Compact, Per Packet Flow Records

- Provides richness and scalability for large networked systems
- Flow Record Key 5-Tuple # of Total # of List of Src IP Src Port Packets **Bytes** Packets Protocol Dest IP Dest Port Flow Packet 1 Flow Packet 2 Flow Packet N Timestamp # Bytes Features Timestamp # Bytes Features Timestamp # Bytes Features IP IP IP TTL TTL TTL Flags Flags Flags PFE Flow Record Overview and Features

 Tailored to fit a user's specific application

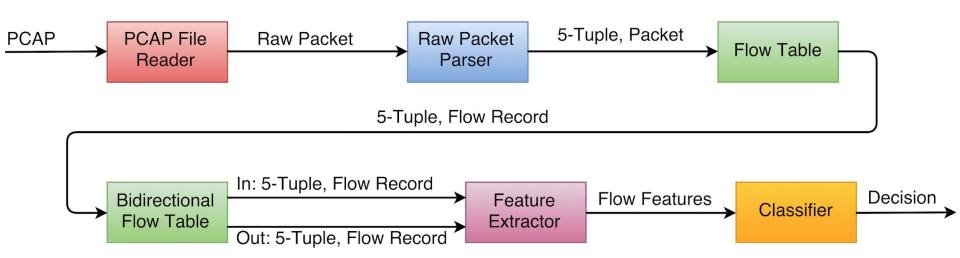


Method

- Goal: Utilize machine learning and leverage the recent trend in switch hardware to identify ransomware via its network traffic signature
- Collect ransomware PCAP samples (>100MB)
- Collect clean traffic as baseline
 - Web browsing, streaming, file downloading, etc.
- Stream processor development
- Classification



Flow Records and Stream Processing





Flow-Based Features

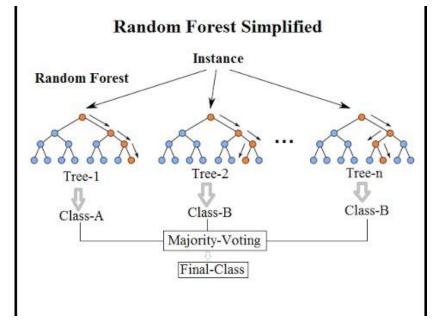
- Flow Duration
- Interarrival Times
 - Minimum
 - Mean
 - Maximum
 - Standard deviation
- Packet Lengths
 - Minimum
 - Mean
 - Maximum
 - Standard deviation

- Burst Lengths
 - Minimum
 - Mean
 - Maximum
- Total number of packets
- Ratios:
 - Packets out/packet in
 - Bytes out/bytes in
- # of unique packet lengths



Random Forest

- Ensemble Algorithm
 - \circ $\,$ Divide and conquer approach
- Collection of decision trees
 - Avoids overfitting
- Random subsets of features used to build smaller, shallower trees
- Majority voting from decision trees to decide class
- Bagging used to improve stability, reduce variance, and increase accuracy



Source: https://www.youtube.com/watch?v=D_2LkhMJcfY

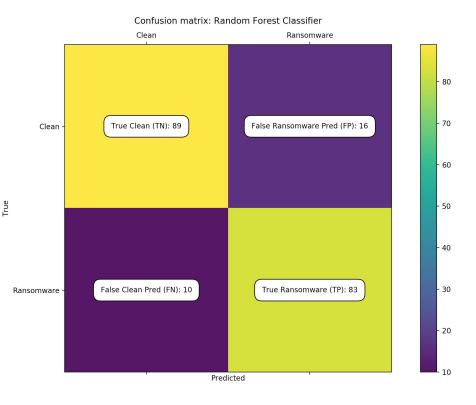


Results



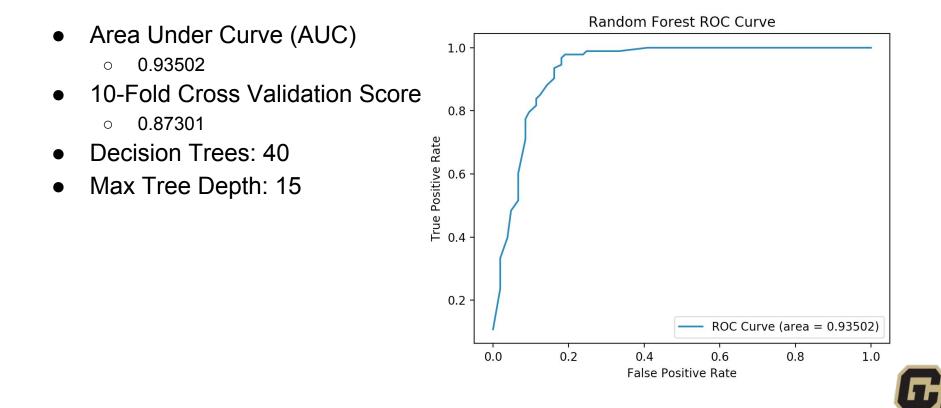
Confusion Matrix (28 Features)

- Accuracy
 - Correct / total = **0.8689**
- Recall
 - tp / (tp + fn) = **0.8925**
- Precision
 - tp / (tp + fp) = **0.8384**
- F1 Score
 - 2 * (R * P) / (R + P) = **0.8689**

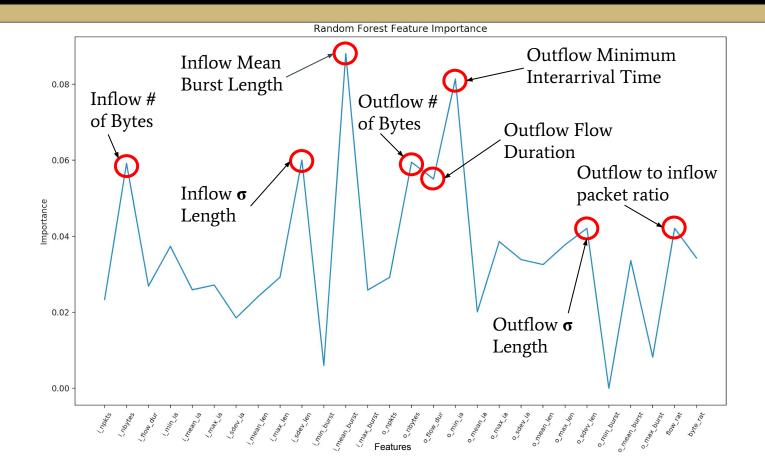




ROC Curve (28 Features)



Feature Importances (28 Features)

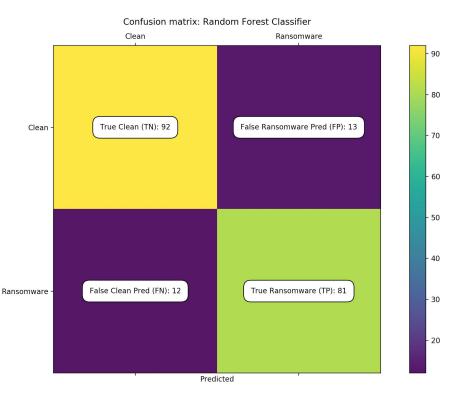




Confusion Matrix (8 Features)

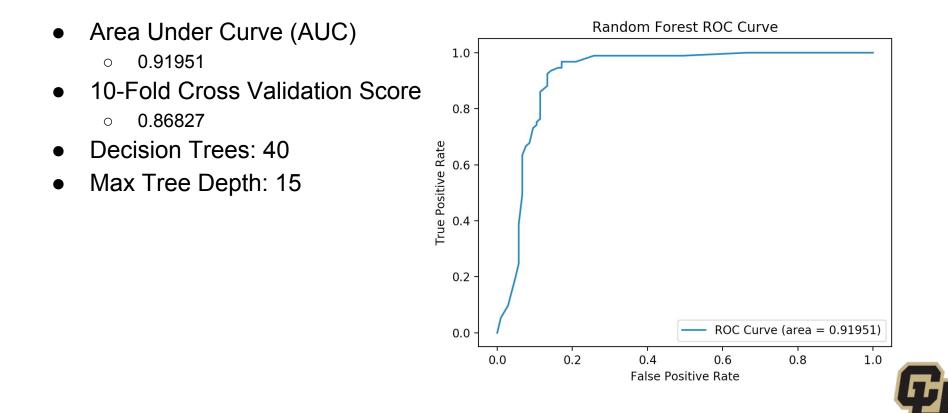
True

- Accuracy
 - Correct / total = **0.8738**
- Recall
 - tp / (tp + fn) = **0.8710**
- Precision
 - tp / (tp + fp) = **0.8617**
- F1 Score
 - 2 * (R * P) / (R + P) = **0.8738**

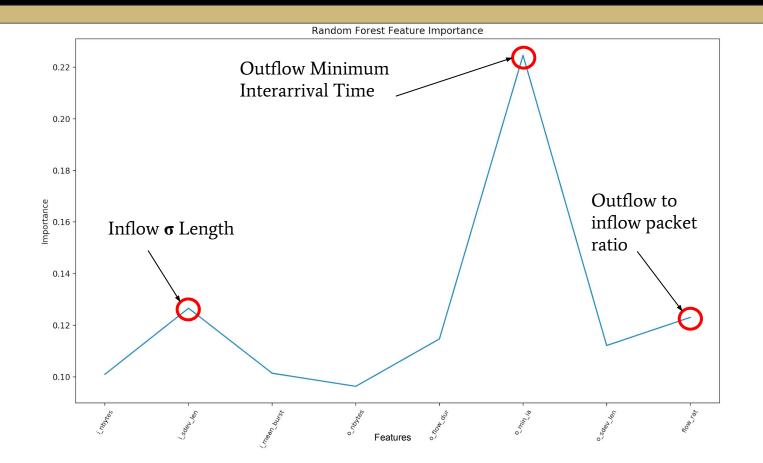




ROC Curve (8 Features)



Feature Importances (8 Features)



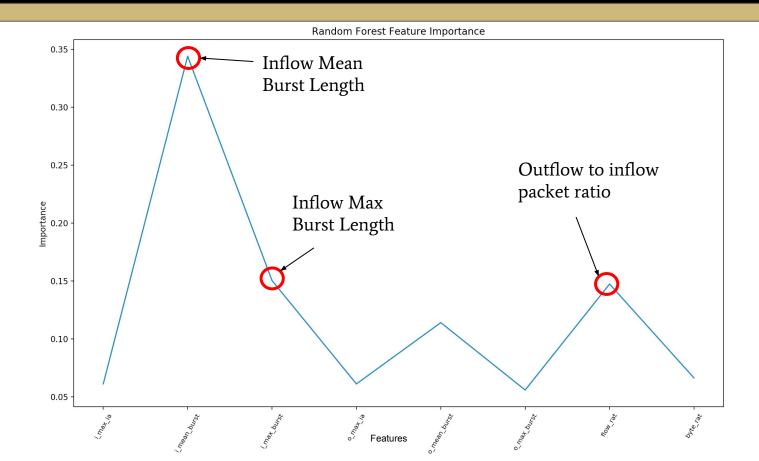


Crypto-Based Cerber Ransomware Detection

Can we identify a specific type of ransomware?



Cerber Feature Importances (8 Features)

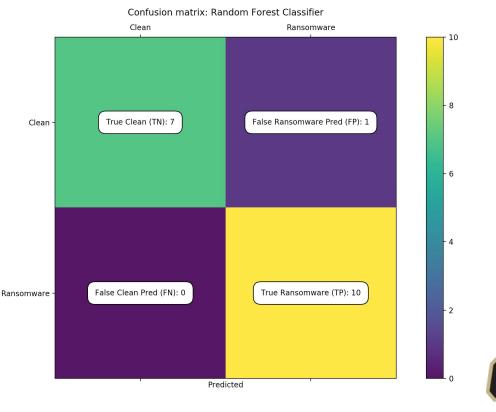




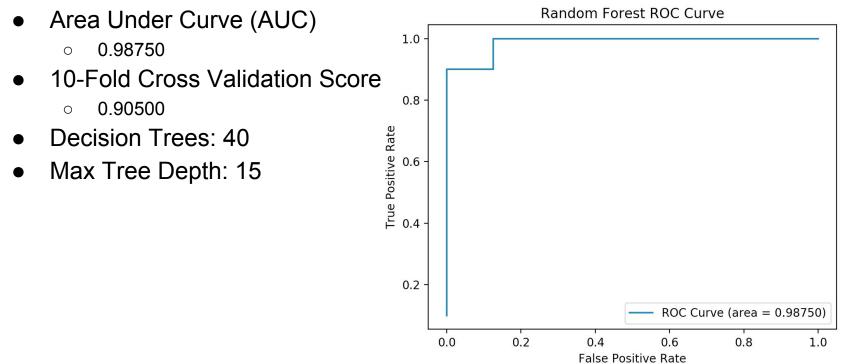
Confusion Matrix Cerber Ransomware

Irue

- Accuracy
 - Correct / total = **0.9444**
- Recall
 - tp / (tp + fn) = **1.000**
- Precision
 - tp / (tp + fp) = **0.9091**
- F1 Score
 - 2 * (R * P) / (R + P) = **0.9439**



ROC Curve Cerber Ransomware







• Initial findings are promising but require further research

• Packet lengths, interarrival times, and flow ratios leave ransomware susceptible to identification

• Recent emergence of PFEs provide the right backbone for flow-based feature extraction



Work in Progress

• Sandboxing ransomware samples to collect network traffic

• Implementing stream processor on a PFE ASIC

• Developing LSTM Recurrent Neural Network

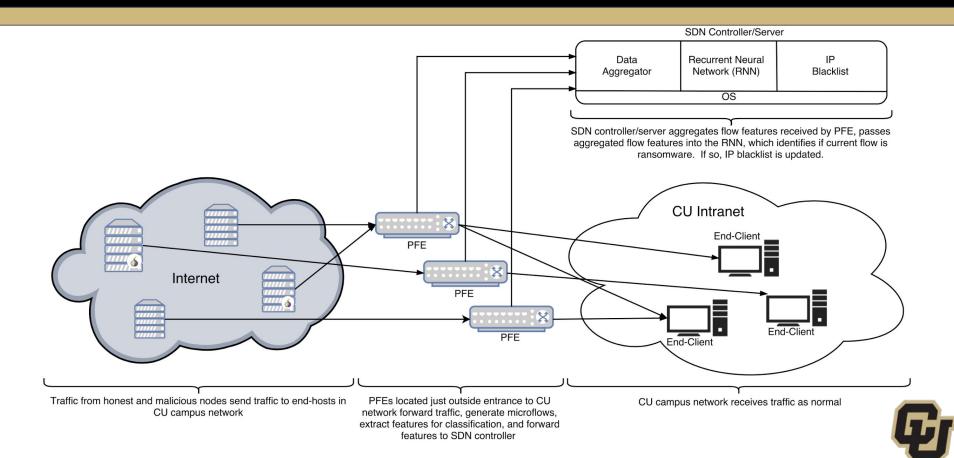
• System architecture redesign



Questions?

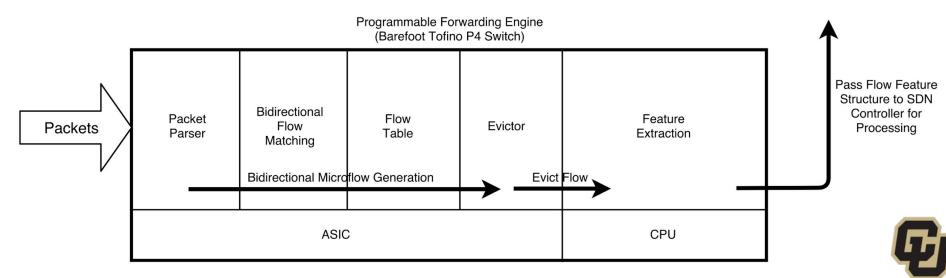


Ransomware Detection System Overview

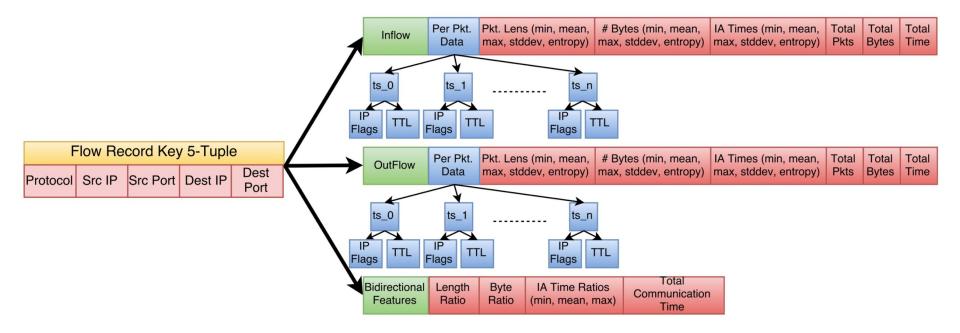


Hardware-generated, Bidirectional Microflows

- Generated in switch ASIC
- Turn current *Flow flow table into bidirectional flow table
- Evict bidirectional microflows to CPU for feature extraction
- PFE data flow overview:

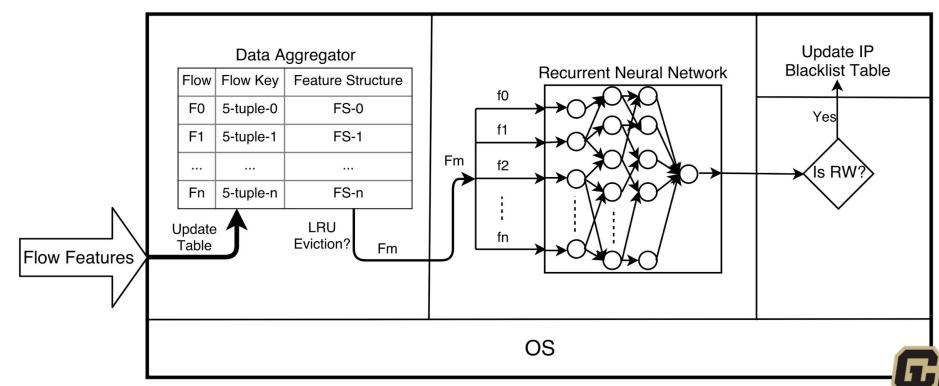


Flow Feature Structure





SDN Controller/Server Data Flow



SDN Controller/Server

Packet Length Frequency

