Using an Ensemble of One-Class SVM Classifiers to Harden Payload-Based Anomaly Detection Systems

Roberto Perdisci[^,+], Guofei Gu[^], Wenke Lee[^]

[^]**Georgia Institute of Technology**, Atlanta, GA, USA

[+]**University of Cagliari**, ITALY

presented by **Roberto Perdisci**

Outline

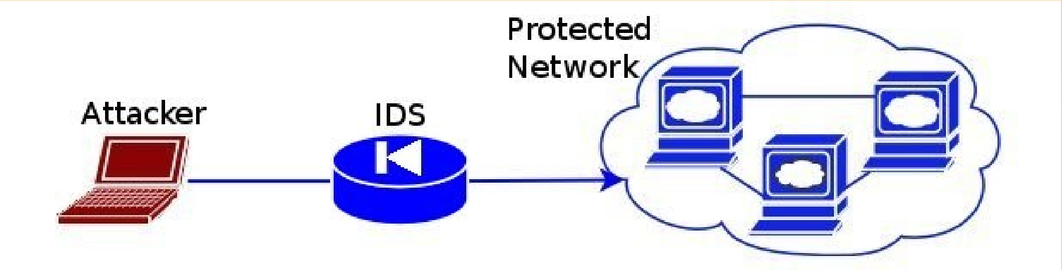
- Anomaly Detection in Computer Networks
- PAYL, a PAYLoad-based Anomaly Detector
- Polymorphic Blending Attack
- Hardening Payload-based Anomaly Detection
  - Payload Analysis using 2ν-grams
  - Combining Multiple One-Class Classifiers
- Experimental Results
- Conclusion

Outline

- Anomaly Detection in Computer Networks
- PAYL, a PAYLoad-based Anomaly Detector
- Polymorphic Blending Attack
- Hardening Payload-based Anomaly Detection
  - Payload Analysis using 2ν-grams
  - Combining Multiple One-Class Classifiers
- Experimental Results
- Conclusion

Anomaly Detection in Computer Networks

- Problem Definition
  - Classify computer network traffic
  - Distinguish between *normal* traffic and *attacks*
  - No labelled dataset



- Assumptions
  - The vast majority of the network traffic is normal
  - Network attacks can be distinguished from normal traffic using suitable metrics

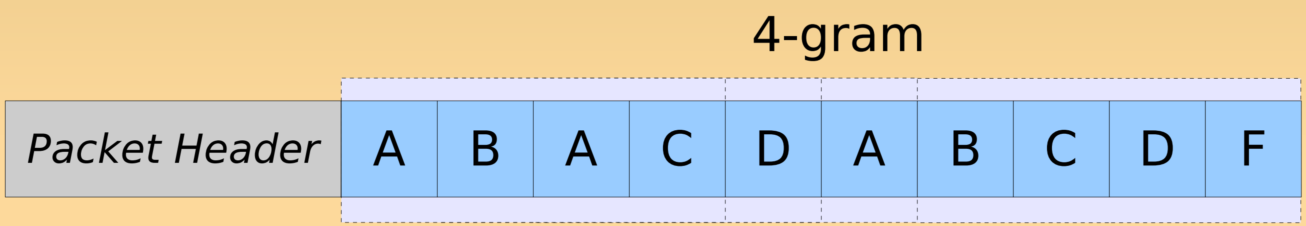
- <u>Outlier Detection problem</u>

Outline

- Anomaly Detection in Computer Networks
- PAYL, a PAYLoad-based Anomaly Detector
- Polymorphic Blending Attack
- Hardening Payload-based Anomaly Detection
  - Payload Analysis using 2ν-grams
  - Combining Multiple One-Class Classifiers
- Experimental Results
- Conclusion

PAYL

- PAYLoad-based Anomaly Detector
  - Developed at Columbia University, NY
  - Based on occurrence frequency of n-grams (sequences of n bytes) in the payload



- Training
  - Frequency of n-grams is extracted for each payload in a (noisy) dataset of *normal* traffic
  - A simple model is constructed by computing the average and standard deviation of frequency of n-grams
  - 256^n possible n-grams = 256^n features

PAYL

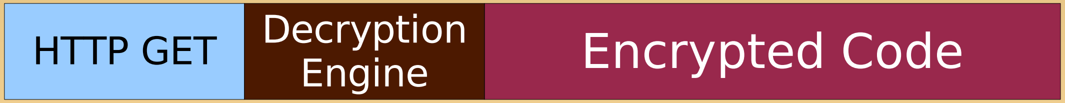
- Operational Phase
  - The frequency of n-grams is extracted from the payload of each packet entering the network
  - *Simplified* Mahalanobis distance used to compare the packet under test to the model of normal traffic
  - An alarm is flagged if distance greater than a certain threshold
- Problems
  - PAYL assumes there is no correlation among features
  - Uses 1-gram (or 2-gram) analysis because high values of *n* are impractical
    - if *n is high* -> curse of dimensionality
    - if *n is low* -> low amount of structural information

Outline

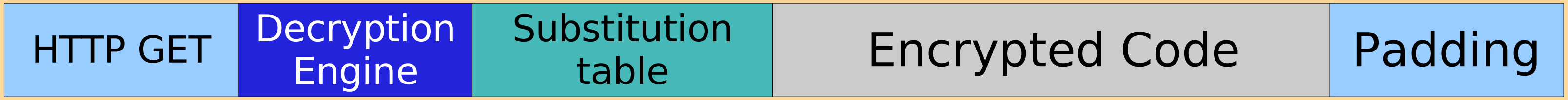
- Anomaly Detection in Computer Networks
- PAYL, a PAYLoad-based Anomaly Detector
- Polymorphic Blending Attack
- Hardening Payload-based Anomaly Detection
  - Payload Analysis using 2ν-grams
  - Combining Multiple One-Class Classifiers
- Experimental Results
- Conclusion

Polymorphic Blending Attack

- Polymorphism is used by attackers to avoid signature-based detection

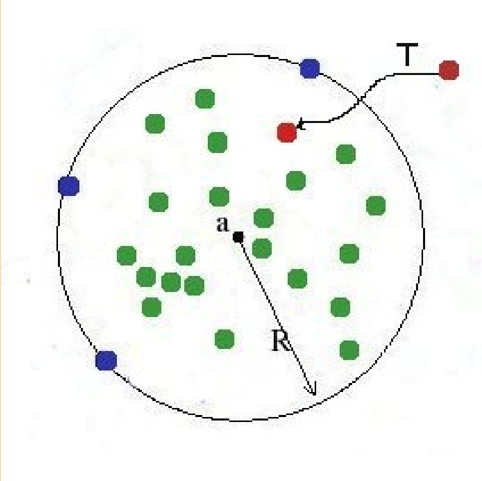


- 1-gram and 2-gram PAYL can easily detect “standard” and Polymorphic attacks
  - normal HTTP requests are highly structured, they contain mostly printable characters
  - the Executable Code, the Decryption Engine and the Encrypted Code contain lots of “unusual” characters (e.g., non-printable)
- Polymorphic Blending Attack can *evade* PAYL
  - Encryption algorithm is designed to make the attack **look like normal traffic**



Polymorphic Blending Attack

- Attack strategy
  - Estimate frequency distribution of n-grams in normal traffic (e.g., sniffing traffic sent towards the victim network)
  - Encode the attack payload to approximate the learned distribution
  - Add padding bytes to further adjust the distribution of n-grams in the attack payload



- Can evade 1-gram and 2-gram PAYL
  - Attack transformation *T* brings the attack pattern inside the decision surface

Analysis of Polymorphic Blending Attack

- Why does the Blending Attack work?
  - Model of normal traffic constructed by PAYL is too simple
  - 1-gram and 2-gram analysis do not extract enough *structural information*
- Shortcomings of the attack
  - Polymorphic Blending Attack uses a greedy algorithm to find a sub-optimal attack transformation
  - The attack transformation is less and less likely to find a good solution for high values of *n*

Outline

- Anomaly Detection in Computer Networks
- PAYL, a PAYLoad-based Anomaly Detector
- Polymorphic Blending Attack
- Hardening Payload-based Anomaly Detection
  - Payload Analysis using 2ν-grams
  - Combining Multiple One-Class Classifiers
- Experimental Results
- Conclusion

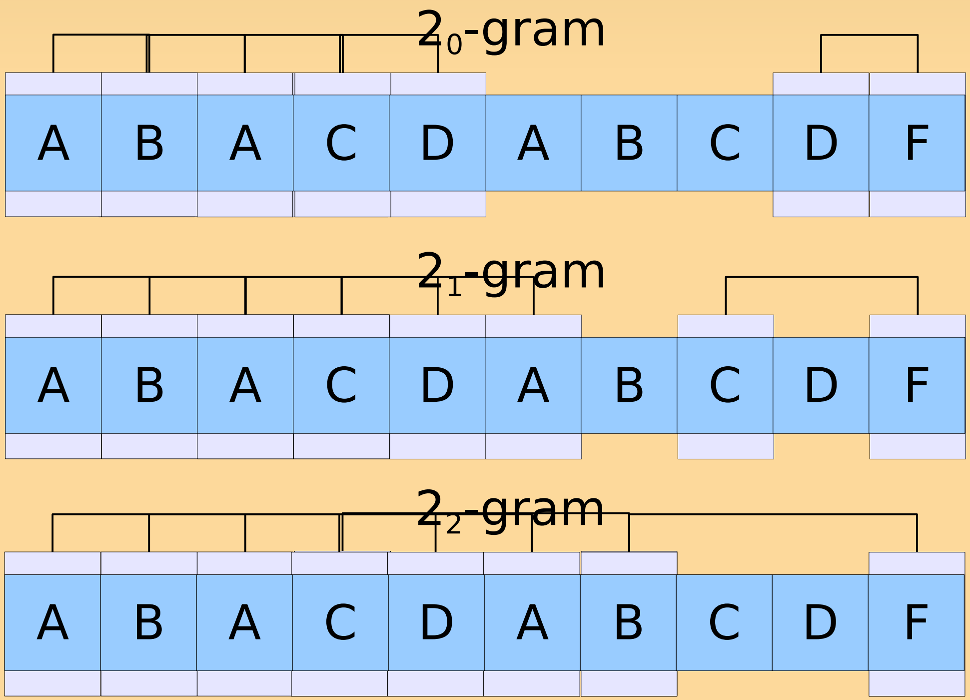
Extracting structural information

- We could use n-gram analysis with a high value of *n*, but...
  - 256^n features! (if n=3 we have 16,777,216 features!)
  - curse of dimensionality
  - problems related to computational cost and memory consumption of learning algorithms

- Observation
  - if n=2 we have 256^2=65,536 features
  - in this case the classification problem is still tractable

2ν-gram analysis

- Definition
  - 2ν-gram = 2 bytes in the payload that are ν bytes apart from each other
  - instead of measuring the occurrence frequency of n-grams we measure the freq. of 2ν-grams, with ν=0..(n-2)



Combining multiple models

- Intuition
  - combining the structural information extracted using the 2ν-gram analysis, ν=0..(n-2) approximately reconstructs the structural information extracted by n-gram analysis
- In practice
  - using 2ν-gram analysis we obtain (n-2+1) different descriptions of the payload
  - each description projects the payload in a 256^2-dimensional feature space
  - construct one model of normal traffic for each value of ν=0..(n-2) using *One-Class SVM*
  - combine the output of the obtained (n-2+1) classifiers using the Majority Voting combination rule

Feature Reduction

- $256^2 = 65{,}536$ features!
  - we **need to reduce the dimensionality** of each of the (n-2+1) feature spaces before constructing classifiers
- **Payload-based Anomaly Detection** with n-gram analysis **is analogous to text classification**
  - true if we consider the bag-of-words technique with freq. of words as features
  - **n-grams = words**
  - **payload = document**
- We use a **Feature Clustering** algorithm proposed for text classification problems
  - Dhillon et al., "*A divisive information-theoretic feature clustering algorithm for text classification*", JMLR 2003

Summary

- Our approach to make Polymorphic Blending Attack harder to succeed
  - Extract more structural information from the payload
  - Construct descriptions of the payload in different feature spaces
  - Reduce the dimensionality of these feature spaces
  - Construct a One-Class SVM classifier on each of the reduced feature spaces to model normal traffic
  - Combine the output of the constructed classifiers

Outline

- Anomaly Detection in Computer Networks
- PAYL, a PAYLoad-based Anomaly Detector
- Polymorphic Blending Attack
- Hardening Payload-based Anomaly Detection
  - Payload Analysis using 2ν-grams
  - Combining Multiple One-Class Classifiers
- Experimental Results
- Conclusion

Experimental Results

- Datasets
  - HTTP requests towards www.cc.gatech.edu collected between October and November 2004
  - Training dataset
    - 1 day of normal traffic (384,389 payloads)
  - Test datasets
    - 4 days of normal traffic (1,315,433 payloads)
  - Attack Dataset (126 payloads)
    - 11 non-polymorphic Buffer Overflow attacks
    - 6 polymorphic attacks
    - 1 Polymorphic Blending Attack (trained to evade 1-gram and 2-gram PAYL)

Experimental Results

1-gram PAYL

DFP(%)	RFP(%)	Detected attacks	DR(%)
0.0	0.00022	1	0.8
0.01	0.01451	4	17.5
0.1	0.15275	17	69.1
1.0	0.92694	17	72.2
2.0	1.86263	17	72.2
5.0	5.69681	18	73.8
10.0	11.05049	18	78.6

2-gram PAYL

DFP(%)	RFP(%)	Detected attacks	DR(%)
0.0	0.00030	14	35.2
0.01	0.01794	17	96.0
0.1	0.12749	17	96.0
1.0	1.22697	17	97.6
2.0	2.89867	17	97.6
5.0	6.46069	17	97.6
10.0	11.25515	17	97.6

Multiple One-Class SVM (n=12,k=40)

DFP(%)	RFP(%)	Detected attacks	DR(%)
0.0	0.0	0	0
0.01	0.00381	17	68.5
0.1	0.07460	17	79.0
1.0	0.49102	18	99.2
2.0	1.14952	18	99.2
5.0	3.47902	18	99.2
10.0	7.50843	18	100

DFP = False positives on **training dataset**

RFP = False positives on **test dataset**

DR = Percentage of **detected attack packets**

Outline

- Anomaly Detection in Computer Networks
- PAYL, a PAYLoad-based Anomaly Detector
- Polymorphic Blending Attack
- Hardening Payload-based Anomaly Detection
  - Payload Analysis using 2ν-grams
  - Combining Multiple One-Class Classifiers
- Experimental Results
- Conclusion

Conclusion

- We introduced the 2ν-gram analysis technique to extract information from the payload
- We used the analogy between payload-based anomaly detection and text classification for feature reduction
- We used an ensemble of classifiers to “combine” the structural information extracted with the 2ν-gram technique
- This makes the Polymorphic Blending Attack more difficult to succeed

Related Work

- **Wang** et al. "*Anomalous Payload-based Network Intrusion Detection*". RAID 2004.
- **Fogla** et al. "*Polymorphic Blending Attack*". USENIX Security 2006.
- **Dhillon** et al. "*A divisive information-theoretic feature clustering algorithm for text classification*", MIT Journal of Machine Learning Research, Vol. 3, 2003
- **Barreno** et al. "*Can machine learning be secure?*". AsiaCCS'06.

Anomaly vs. Signature-based Detection

- Signature-based IDS are the most deployed
  - efficient patter matching
  - can detect known attacks
  - low number of false positives (i.e., false alarms)
  - **not able to detect unknown** (zero-day) **attacks**
- Anomaly Detection
  - can **detect known and unknown attacks** (in theory!)
  - difficulties in precisely modelling the normal traffic
  - may generate a higher number of false positives compared to signature-based IDS

Polymorphic Attack

- A “standard” Buffer Overflow attack (for example) looks like

  – these attacks can usually be detected using pattern matching (signature-based IDS)

- Polymorphism is used by attackers to avoid signature-based detection

  – the Decryption Engine and the Encrypted Code change every time the attack is launched towards a new victim

Experimental Results

- Single One-Class SVM classifiers
  - *RBF* kernel (γ=0.5)
  - *k* = number of Feature Clusters
  - ν = parameter for the 2ν-gram analysis

ν	*k* = 10	20	40	80	160
0	0.9660 (0.4180E-3)	0.9664 (0.3855E-3)	0.9665 (0.4335E-3)	0.9662 (0.2100E-3)	**0.9668** (0.4686E-3)
1	0.9842 (0.6431E-3)	0.9839 (0.7047E-3)	**0.9845** (0.7049E-3)	0.9833 (1.2533E-3)	0.9837 (0.9437E-3)
2	0.9866 (0.7615E-3)	0.9867 (0.6465E-3)	0.9875 (0.6665E-3)	**0.9887** (2.6859E-3)	0.9862 (0.7753E-3)
3	0.9844 (1.2207E-3)	0.9836 (1.1577E-3)	**0.9874** (1.0251E-3)	0.9832 (1.0619E-3)	0.9825 (0.6835E-3)
4	0.9846 (0.5612E-3)	0.9847 (1.5334E-3)	0.9846 (0.9229E-3)	0.9849 (1.5966E-3)	**0.9855** (0.4649E-3)
5	0.9806 (0.8638E-3)	0.9813 (0.9072E-3)	0.9810 (0.5590E-3)	0.9813 (0.8494E-3)	**0.9818** (0.3778E-3)
6	0.9809 (0.7836E-3)	0.9806 (1.1608E-3)	**0.9812** (1.6199E-3)	0.9794 (0.3323E-3)	0.9796 (0.4240E-3)
7	0.9819 (1.6897E-3)	0.9854 (0.8485E-3)	0.9844 (1.2407E-3)	0.9863 (1.9233E-3)	**0.9877** (0.7670E-3)
8	0.9779 (1.7626E-3)	0.9782 (1.9797E-3)	0.9787 (2.0032E-3)	**0.9793** (1.0847E-3)	0.9785 (1.7024E-3)
9	0.9733 (3.1948E-3)	**0.9775** (1.9651E-3)	0.9770 (1.0803E-3)	0.9743 (2.4879E-3)	0.9722 (1.2258E-3)
10	0.9549 (2.7850E-3)	0.9587 (3.3831E-3)	0.9597 (3.8900E-3)	0.9608 (1.2084E-3)	**0.9681** (7.1185E-3)

AUC measured in the interval [0,0.1] of false positives (normalized)

Advantages of our approach

- The attacker could evade our IDS if he was able to construct the attack transformation to approximate the distribution of (n/2+1)-grams in normal traffic
- However, the greedy attack transformation algorithm is unlikely to find a good solution if (n/2+1) is a sufficiently high value
- A new attack transformation algorithm specifically crafted to approximate the distribution of 2ν-grams has to evade at least n/2 different models at the same time
- The introduced overhead added to the operational phase is expected to be fairly low