CS259D: Data Mining for Cybersecurity

One-class classification

- Most samples from target class
- Rejection rate
  - % of training data points classified as outliers
  - Allows for presence of noise
  - Tolerable false positive rate

One-class SVM

- Hyperplane separating training samples from the feature space origin
  - May not always exist in original feature space
  - Feature space mapped to a Kernel space
  - With Gaussian kernel, hyperplane always exists

$$K(x,y) = \Phi(x).\Phi(y) = \exp(-\frac{\| x - y \|^2}{2s})$$

$$\min_{w,\xi,\rho}(\frac{1}{2}\| w \|^2 - \rho + \frac{1}{hC}\sum_{i=1}^{h}\xi_i)$$

$$w.\Phi(xi) \geq \rho - \xi i \qquad (1 \leq i \leq h)$$

$$f_{svc}(z) = I(\sum_i \alpha_i K(x_i,z) \geq \rho); \quad \sum_{i=1}^{h}\alpha_i = 1$$

Class-conditional probability formulation

- Conditional probability representation:

$$p(x \mid w_t) = \frac{1}{(2\pi s)^{d/2}} \sum_{i=1}^{n} \alpha_i K(x, x_i)$$

- This is actually a distribution
- Classify as normal if:

$$p(x \mid w_t) \geq \rho'; \quad \rho' = \rho / (2\pi s)^{d/2}$$

Fusion rules

- Min, Max, Mean, Product
- Applied to a-posteriori class probabilities under different models: $P_i(w_j \mid x)$
- Assuming uniform distribution for outliers can turn these rules into class-conditional probabilities

Combining one-class SVM classifiers

- Average:

$$y_{avg}(x) = \frac{1}{L}\sum_{i=1}^{L} p_i(x \mid w_t)$$

$$y_{avg}(x) < \theta \Rightarrow outlier$$

- Product:

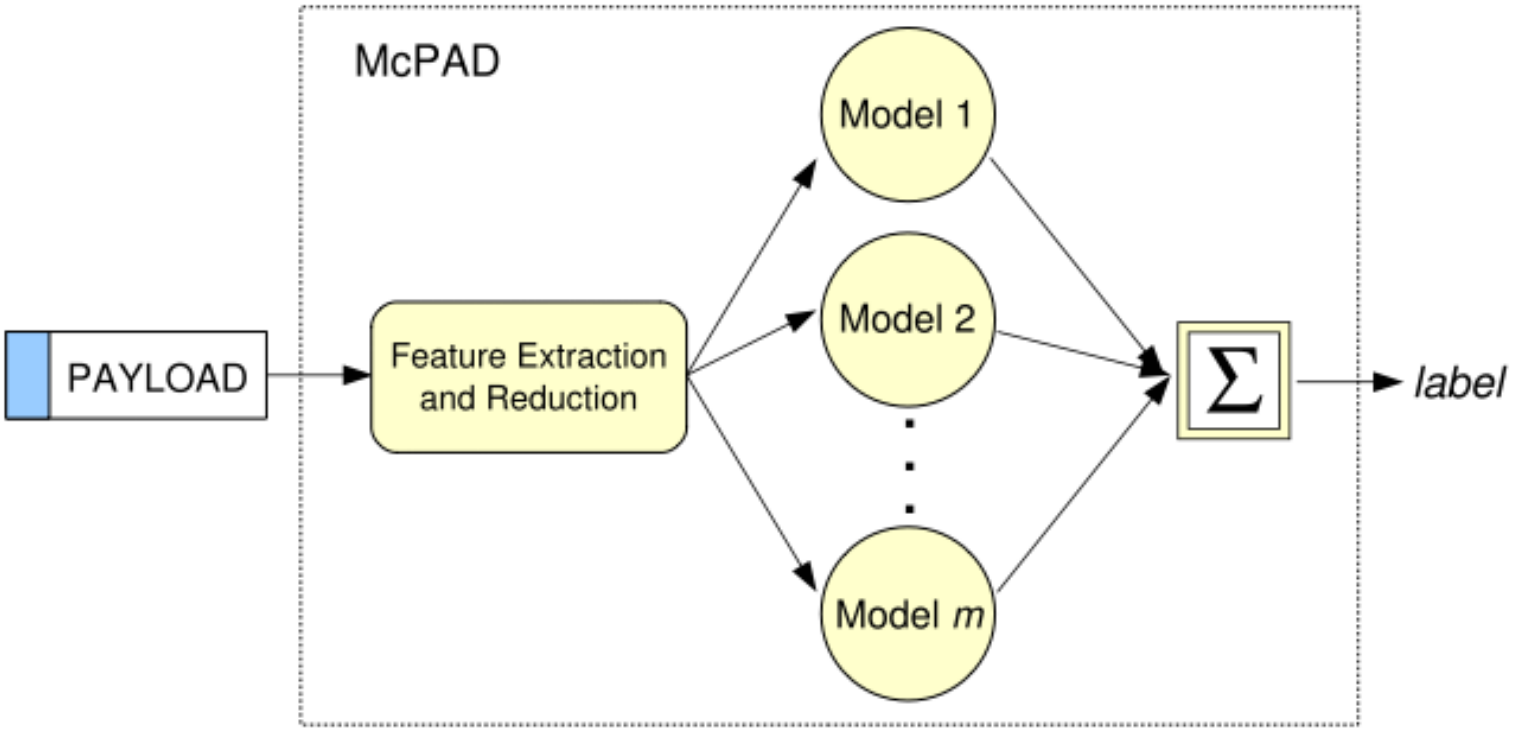
$$y_{prod}(x) = \prod_{i=1}^{L} p_i(x \mid w_t)$$

- Min and Max rules similarly defined
- Majority voting rule: Class voted for by majority of classifiers

McPAD: Multiple Classifier System for Accurate Payload-based Anomaly Detection

- Features (2_v-grams): frequencies of bytes that are v bytes apart
- 256^2 features irrespective of v
- Computed using a sliding window of size v +2
  - Marginalized distribution
- v=0 is the 2-gram model of PAYL
- Combination to reconstruct sequence information
- Feature clustering to reduce dimension

McPAD architecture



Experiments: Parameters

γ	0.5
ν	0-10
Feature Clusters (k)	10, 20, 40, 80, 160
Desired FP Rate	10%, 5%, 2%, 1%, 0.5%, 0.2% 0.1%, 0.05%, 0.02%, 0.01%, 0.001%

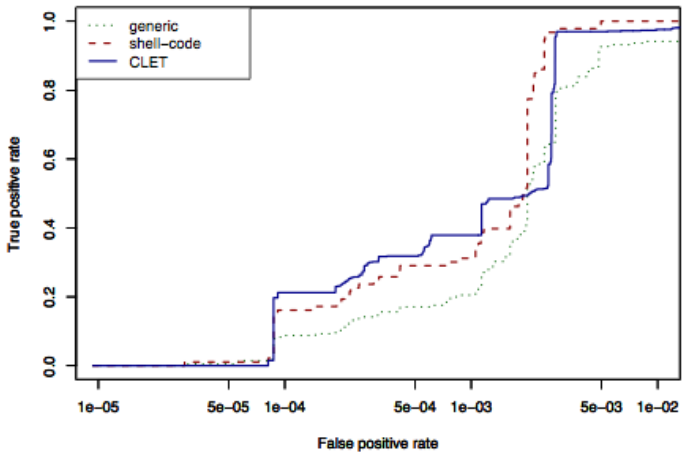
Attacks

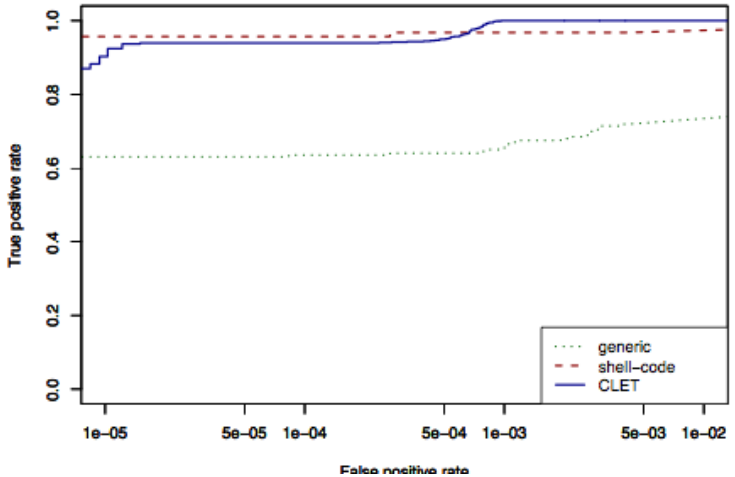
- Generic attacks
- Shell-code attacks
- CLET attacks
- PBA attacks

Results

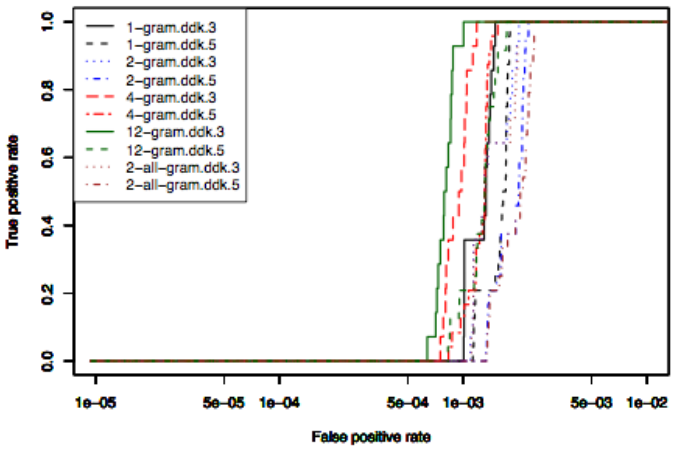
	Maj. Vot.	Avg. Prob.	Prod. Prob.	Min. Prob.	Max. Prob.
Generic Attacks					
k=10	0.83501	0.86331	0.8633	0.87187	0.76765
k=20	0.8366	0.8613	0.86135	0.86882	0.7492
k=40	0.8366	0.86312	0.86407	0.87783	0.77834
k=80	0.84778	0.85948	0.8595	0.88594	0.80212
k=160	0.87016	0.8884	0.88828	0.87131	0.69164
Shell-code Attacks					
k=10	0.98632	0.99544	0.99543	0.99323	0.94105
k=20	0.98758	0.99689	0.9969	0.99361	0.94685
k=40	0.98903	0.99827	0.99826	0.99417	0.97585
k=80	0.99613	0.99874	0.99875	0.9965	0.98666
k=160	0.98723	0.99785	0.99775	0.99709	0.76661
CLET Attacks					
k=10	0.99776	0.99854	0.99854	0.99866	0.9589
k=20	0.99778	0.99839	0.99839	0.99925	0.969
k=40	0.99757	0.99815	0.99815	0.99908	0.98624
k=80	0.99773	0.99785	0.9979	0.99925	0.99669
k=160	0.99737	0.9985	0.99844	0.99913	0.83275

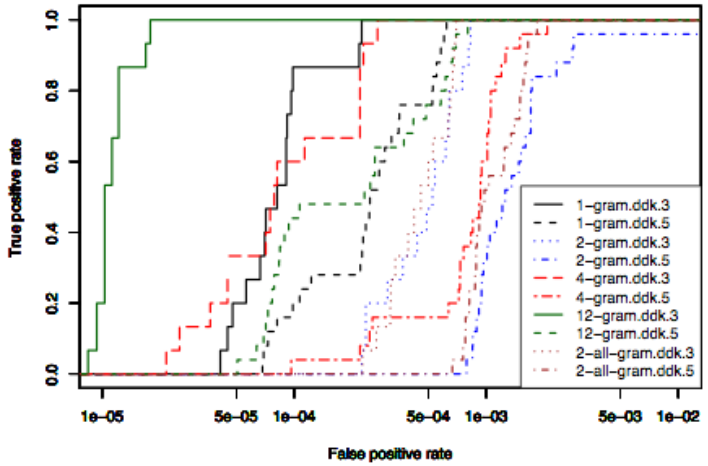
Results


generic
shell−code
CLET
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
1e−05
5e−05
1e−04
5e−04
1e−03
5e−03
1e−02


generic
shell−code
CLET
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
1e−05
5e−05
1e−04
5e−04
1e−03
5e−03
1e−02

Results


1−gram.ddk.3
1−gram.ddk.5
2−gram.ddk.3
2−gram.ddk.5
4−gram.ddk.3
4−gram.ddk.5
12−gram.ddk.3
12−gram.ddk.5
2−all−gram.ddk.3
2−all−gram.ddk.5
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
1e−05
5e−05
1e−04
5e−04
1e−03
5e−03
1e−02


1−gram.ddk.3
1−gram.ddk.5
2−gram.ddk.3
2−gram.ddk.5
4−gram.ddk.3
4−gram.ddk.5
12−gram.ddk.3
12−gram.ddk.5
2−all−gram.ddk.3
2−all−gram.ddk.5
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
1e−05
5e−05
1e−04
5e−04
1e−03
5e−03
1e−02

References

- "McPAD: A Multiple Classifier System for Accurate Payload-Based Anomaly Detection", Perdisci et al, 2009