System Configuration and Security

Gaspar Modelo-Howard

Paper

- Cárdenas, A., Baras, J., & Seamon, K. ***A framework for the evaluation of intrusion detection systems.*** Proc. of the 2006 IEEE Symposium on Security & Privacy (S&P'06)

Paper's Q&A

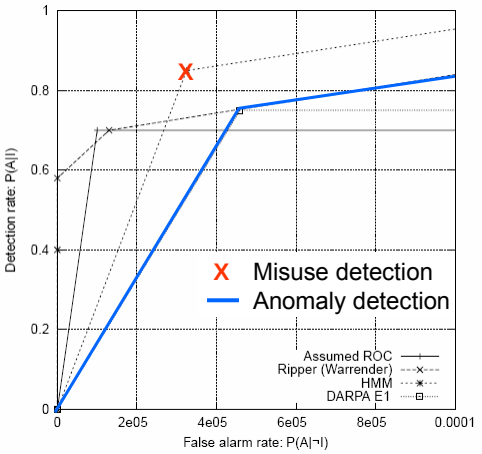
- Motivation
  - How to compare two or more IDS?
  - ***How to evaluate the performance of an IDS?***
  - ***How to determine the best configuration of an IDS?***
- Synopsis
  - Comparison of IDS performance metrics
  - IDOC curve
  - Introduction of formal framework for reasoning about IDS performance
  - Proposed metrics against adaptive adversaries

Agenda

- Notation and Definitions
- Evaluation Metrics
- Graphical Analysis
- Threat Models and Security Properties of the Evaluation
- Conclusions

Notation and Definitions

- IDS is an algorithm that receives event fatures *X* = {x[1], x[2],…,} and classifies each input as being normal or attack
  - *IDS*: *X* → {A,¬A}
- Detection rate (DR)
  - P_D=*Pr[A=1|I=1]*
- False alarm rate (FAR)
  - P_{FA}=*Pr[A=1|I=0]*
- Positive Predictive Value (PPV)
  - *PPV=Pr[I=1|A=1]*
- Base rate (BR)
  - *p=Pr[I=1]*



ROC Curve

Evaluation Metrics

- Expected Cost
  - Besides DR and FAR, can depend on other factors (hostility of environment, IDS operational costs, expected damage done by security breaches, etc.)

State of the system	Detector's report	
	No Alarm (A=0)	Alarm (A=1)
No Intrusion ($I=0$)	$C(0,0)$	$C(0,1)$
Intrusion ($A=1$)	$C(1,0)$	$C(1,1)$

  - For an intrusion $R(1,P_D) \equiv C(1,0)(1-P_D)+C(1,1)P_D$
  - For a non-intrusion $R(0,P_{FA}) \equiv C(0,0)(1-P_{FA})+C(0,1)P_{FA}$
  - Overall expected cost $\mathbf{E}[C(I,A)] = R(0,P_{FA})(1-p)+R(1,P_D)p$
  - Problem is to find optimal tradeoff between P_D and P_{FA} such that **E**[*C(I,A)*] is minimized

Evaluation Metrics

- Intrusion Detection Capability (C_{ID})
  - Objective metric, not like **E**[*C(I,A)*]

(#)
$$C_{ID} = \frac{\mathbf{I}(I;A)}{\mathbf{H}(I)}$$
← Mutual information
← Entropy

  - By fine tuning an IDS based on CID, we are finding the operating point that maximizes the uncertainty of whether an arbitrary input event x was generated by intrusion or not
  - Main drawback is that it obscures the intuition that is to be expected when evaluating the performance of an IDS

Evaluation Metrics

- Base-Rate Fallacy and Predictive Value Metrics
  - An important cause for large amount of false alarm generation is enormous difference between amount of normal and intrusion events

$$PPV = \Pr[I = 1 \mid A = 1] = \frac{P_D p}{(P_D - P_{FA})p + P_{FA}} = \frac{(1)(10^{-5})}{(1-0.01)(10^{-5}) + 0.01} = 0.000999$$

  - Need to also look at NPV. Actually, find a trade-off between PPV and NPV (want to maximize both)

$$NPV = \Pr[I = 0 \mid A = 0] = \frac{(1-p)(1-P_{FA})}{(1-P_D)p + (1-p)(1-P_{FA})}$$

Graphical Analysis

- Minimization Approach – Expected Cost
- Tradeoff Approach – Intrusion Detector Operating Characteristic (IDOC)

Expected Cost – Minimization Approach

- Desirable to tune uncertain IDS parameters based on feedback from system
  - ROC + Isolines
- ROC curves illustrate behavior of classifier without regard to uncertain parameters
- Isolines = any point on line has equal E[*C(I,A)*]
- Evaluation reduced to finding point of ROC curve that intercepts optimal isoline
  - Limitation: cost specification is a priori and fixed

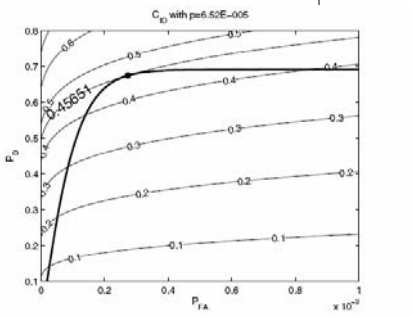


Figure 1. Isoline projections of C_{ID} onto the ROC curve. The optimal C_{ID} value is $C_{ID} = 0.4565$. The associated costs are $C(0,0) = 3 \times 10^{-5}$, $C(0,1) = 0.2156$, $C(1,0) = 15.5255$ and $C(1,1) = 2.8487$. The optimal operating point is $P_{FA} = 2.76 \times 10^{-4}$ and $P_D = 0.6749$.

Expected Cost – Minimization Approach

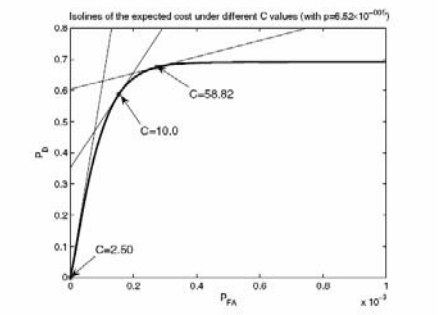


Figure 2. As the cost ratio C increases, the slope of the optimal isoline decreases

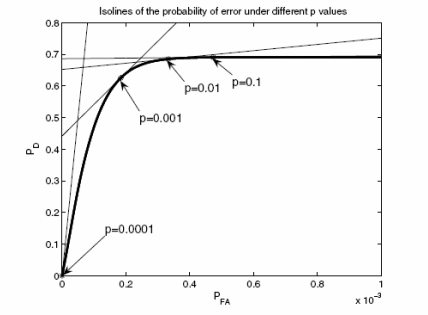


Figure 3. As the base-rate p decreases, the slope of the optimal isoline increases

- Effects of costs
- BR fallacy implications on costs of an IDS

IDOC – Tradeoff Approach

- From previous graphical analysis, intuition for PPV-NPV tradeoff is not clear
  - Use isolines for PPV and NPV

Lemma 1 *Two sets of points* (P_{FA1}, P_{D1}) *and* (P_{FA2}, P_{D2}) *have the same PPV value if and only if*

$$\frac{P_{FA2}}{P_{D2}} = \frac{P_{FA1}}{P_{D1}} = \tan\theta \qquad (11)$$

where θ *is the angle between the line* $P_{FA} = 0$ *and the isoline. Moreover the PPV value of an isoline at angle* θ *is*

$$PPV_{\theta,p} = \frac{p}{p + (1-p)\tan\theta} \qquad (12)$$

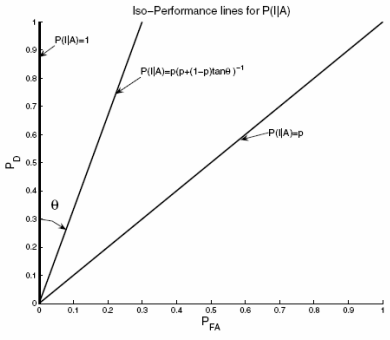


Figure 4. The PPV isolines in the ROC space are straight lines that depend only on θ. The PPV values of interest range from 1 to p

IDOC – Tradeoff Approach

- If *p* is very small → NPV ≈ 1
- Most relevant metrics for performance tradeoff are PPV and P_D
- Should also include *p*

Similarly, two set of points (P_{FA1}, P_{D1}) *and* $(P_{FA2}, PD2)$ *have the same NPV value if and only if*

$$\frac{1-P_{D1}}{1-P_{FA1}} = \frac{1-P_{D2}}{1-P_{FA2}} = \tan\phi \qquad (13)$$

where ϕ *is the angle between the line* $P_D = 1$ *and the isoline. Moreover the NPV value of an isoline at angle* ϕ *is*

$$NPV_{\phi,p} = \frac{1-p}{p(\tan\phi - 1)+1} \qquad (14)$$

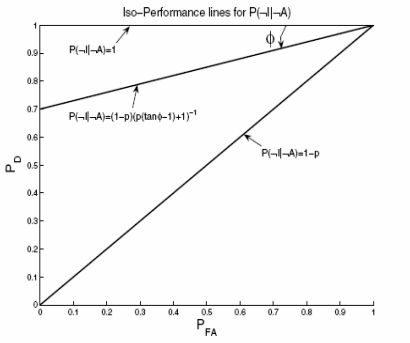


Figure 5. The NPV isolines in the ROC space are straight lines that depend only on ϕ. The NPV values of interest range from 1 to $1-p$

IDOC – Tradeoff Approach

- IDOC shows P_D and *PPV* under different *p*

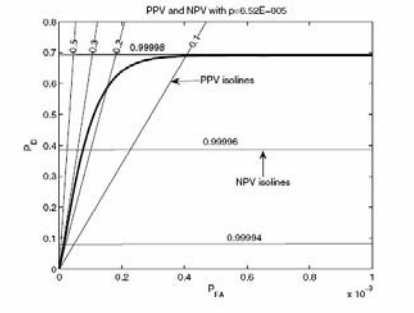


Figure 6. PPV and NPV isolines for the ROC of interest.

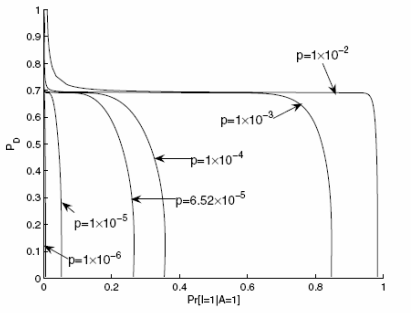


Figure 7. IDOC for the ROC of Figure 6.

Threat Models and Security Properties of the Evaluation

- Formal framework to reason about robustness of IDS evaluation method
- Assumed base-rate $\hat{p}$, operating condition $(\hat{P}_{FA}, \hat{P}_D)$, original *ROC* curve and cost function *C(I,A)* are public values

Definition 1 An $I\mathcal{DS}$ algorithm is the composition of algorithms $\mathcal{D}$ (an algorithm from where we can obtain an ROC curve) and $\mathcal{DM}$ (an algorithm responsible for selecting an operating point). During operation, an $I\mathcal{DS}$ receives a continuous data stream of event features $\mathbf{x}[1], \mathbf{x}[2], \ldots$ and classifies each input $\mathbf{x}[j]$ by raising an alarm or not. Formally:[1]

$I\mathcal{DS}(\mathbf{x})$
$y = \mathcal{D}(\mathbf{x})$
$A \leftarrow \mathcal{DM}(y)$
Output A (where $A \in \{0,1\}$)

Definition 2 A $(\delta, \alpha, \beta) - intruder$ is an algorithm $\mathfrak{I}$ that can select its frequency of intrusions p_1 from the interval $\delta = [\hat{p} - \delta_l, \hat{p} + \delta_u]$. If it decides to attempt an intrusion, then with probability $p_2 \in [0, \beta]$, it creates an attack feature $\mathbf{x}$ that will go undetected by the IDS (otherwise this intrusion is detected with probability $\hat{P}_D$). If it decides not to attempt an intrusion, with probability $p_3 \in [0, \alpha]$ it creates a feature $\mathbf{x}$ that will raise a false alarm in the IDS

$\mathfrak{I}(\delta, \alpha, \beta)$
Select $p_1 \in [\hat{p} - \delta_l, \hat{p} + \delta_u]$
Select $p_2 \in [0, \alpha]$
Select $p_3 \in [0, \beta]$
$I \leftarrow \text{Bernoulli}(p_1)$
If $I = 1$
$B \leftarrow \text{Bernoulli}(p_3)$
$\mathbf{x} \leftarrow \textbf{Feature}(1, (min\{(1-B), \hat{P}_D\}))$
Else
$B \leftarrow \text{Bernoulli}(p_2)$
$\mathbf{x} \leftarrow \textbf{Feature}(0, max\{B, \hat{P}_{FA}\})$
Output $(I, \mathbf{x})$

where Bernoulli(ζ) outputs a Bernoulli random variable with probability of success ζ.

Threat Models and Security Properties of the Evaluation

- Robust Expected Cost Evaluation
  - An evaluation should claim that an IDS is better than others if its expected valued under the worst performance is smaller than the expected value under worst performance of other IDSs

Definition 3 An evaluation method that claims the expected cost of an $I\mathcal{DS}$ is at most r is **robust** against a $(\delta, \alpha, \beta) - intruder$ if the expected cost of $I\mathcal{DS}$ during the attack ($\mathbf{E}^{\delta,\alpha,\beta}[C[I,A]]$) is no larger than r, i.e.

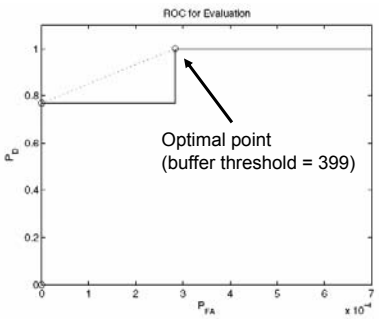
$$\mathbf{E}^{\delta,\alpha,\beta}[C[I,A]] = \sum_{i,a} C(i,a) \times$$
$$\Pr[\,(I,\mathbf{x}) \leftarrow \mathfrak{I}(\delta,\alpha,\beta);\, A \leftarrow I\mathcal{DS}(\mathbf{x}) :\, I = i, A = a\,] \leq r$$

Threat Models and Security Properties of the Evaluation

- Robust Expected Cost Evaluation
  - Using expected cost of IDS under intrusion and no intrusion, determine solution to a zero-sum game between intruder (maximizer) and IDS (minimizer)
  - Basic idea is
    - IF uncertainty range of p is large enough, Nash equilibrium of game is obtained by selecting the point intercepting equation (#)
    - ELSE determine dominant strategy for the intruder, either $\hat{p} + \delta_u$ or $\hat{p} - \delta_l$
- Robust IDOC Evaluation
  - Worst attacker for the evaluation is intruder that selects $p_1 = \hat{p} - \delta_l$, $p_2 = \alpha$ and $p_3 = \beta$

Example: Robust Evaluation of IDSs

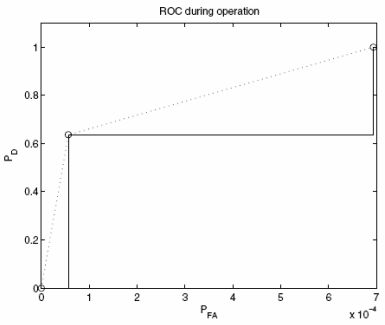
- Used 1998 MIT/Lincoln Labs data set
  - Solaris system log files – buffer overflow detection
  - Monitoring the execution of every program
- Evaluation period
  - IDS performed well
  - p = 13/81108 = 1.6 X 10^{-4}
  - **E**[$C(I,A)$] = 2.83 X 10^{-2}



(a) Original ROC obtained during the evaluation period

Example: Robust Evaluation of IDSs

- Operation period
  - IDS did not performed well
  - **E**[*C(I,A)*] = 6.934 X 10^{-2}
  - Reason is base-rate is smaller during the operation period (p = 7 X 10^{-5})
  - Smaller base-rate should have given us smaller cost

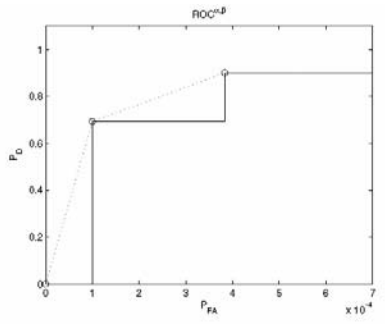


(b) Effective ROC during operation time

Example: Robust Evaluation of IDSs

Let us begin the evaluation process from the scratch by assuming a $([1 \times 10^{-5}, 0], 1 \times 10^{-4}, 0.1) - intruder$, where $\delta = [1 \times 10^{-5}, 0]$ means the IDS evaluator believes that the base-rate during operation will be at most $\hat{p}$ and at least $\hat{p} - 1 \times 10^{-5}$. $\alpha = 1 \times 10^{-5}$ means that the IDS evaluator believes that new normal behavior will have the chance of firing an alarm with probability 1×10^{-5}. And $\beta = 0.1$ means that the IDS operator has estimated that ten percent of the attacks during operation will go undetected.

- Optimal point for curve has **E**[*C(I,A)*] = 5.19 X 10^{-2}
- During operation it had a **E**[*C(I,A)*] = 2.73 X 10^{-2}
- Under traditional evaluation, IDS with buffer threshold length of 399 would have been chosen over IDS of 773 (lower expected cost)



(c) Original ROC under adversarial attack $ROC^{\alpha,\beta}$

Conclusions (from Authors)

- Main problems that any empirical test of an IDS will face
  - Limitations of inference based on experiments alone
  - Evaluating an IDS based on its average performance is not enough since is subject to adversarial environment
- IDOC curve are general method for any classification algorithm whose classes are heavily imbalanced (very small or large *p*).