Measuring Intrusion Detection Capability: An Information-Theoretic Approach

Guofei Gu, Prahlad Fogla, David Dagon, Wenke Lee
Georgia Tech

Boris Skoric
Philips Research Lab

Outline

- Motivation
  - Problem
  - Why existing metrics not enough?
- An Information-Theoretic View of Intrusion Detection
  - Intrusion detection capability: C_{ID}
- Experiment Evaluation
- Conclusion and Future Work

Two Motivating Examples

- Suppose your company is choosing IDS from two candidates
  - IDS1 can detect 10% more attacks, but IDS2 can produce 10% lower false alarms
  - Which one is better?

- Suppose you are configuring your IDS at some operation point (by setting threshold, rule set, policy, …) in your environment
  - How do you set the IDS at an optimal point?

Problem

- A fundamental problem in intrusion detection
  - What metric(s) to objectively measure the effectiveness of an IDS in terms of its ability to correctly classify events as normal or intrusion?
- Why we need such a metric?
  - selecting the best IDS configuration for an operation environment
  - evaluating different IDSs

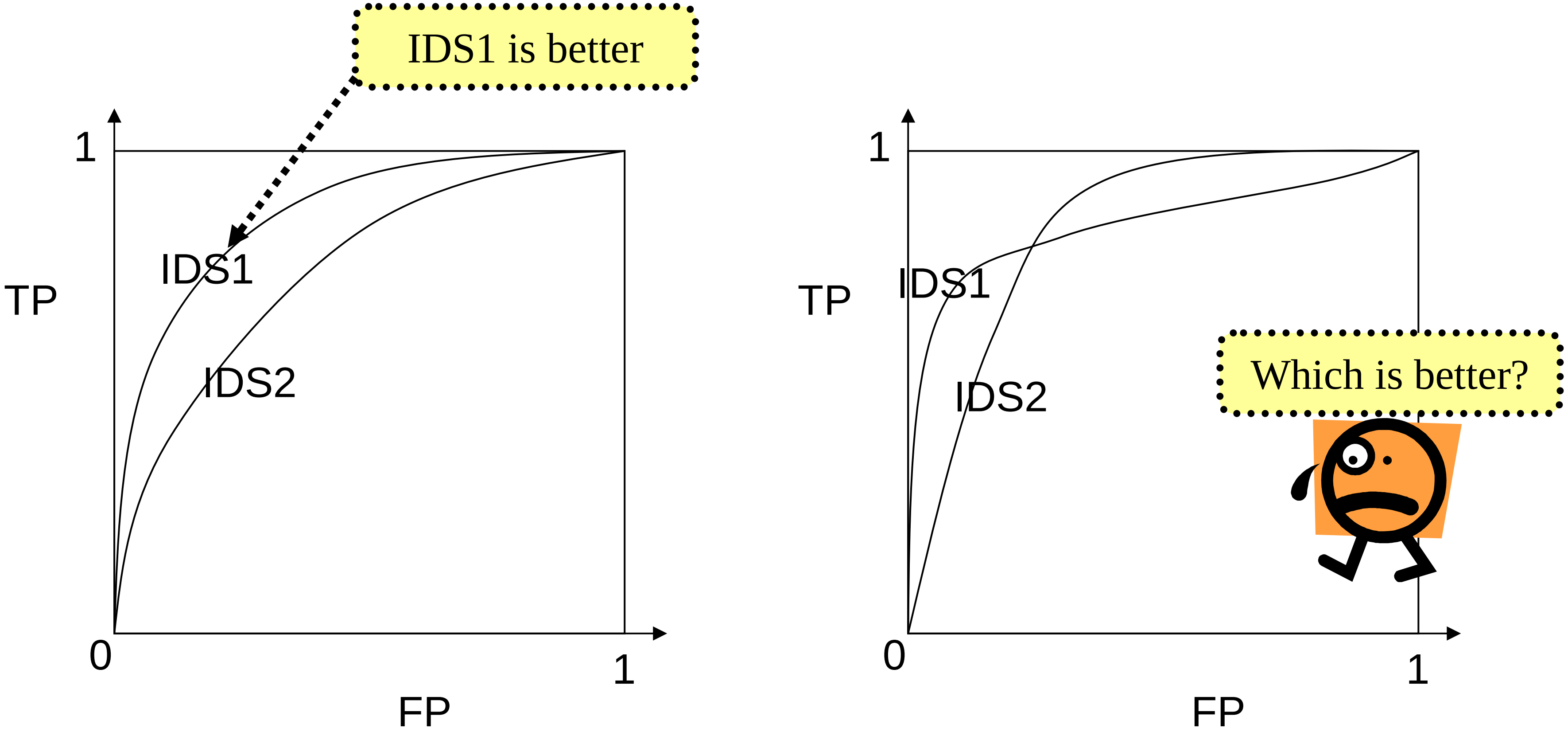
Basic and Commonly Used Metrics

- FP (α): false positive rate
  - $P(A|\neg I)$
- TP ($1-\beta$): true positive rate, or detection rate
  - $P(A|I)$
- Instead of using TP, we can also use FN (β): false negative rate
  - $P(\neg A|I) = 1 - P(A|I)$

Tradeoff is Needed

- Example
  - IDS1: FN=10%, FP=5%
  - IDS2: FN=20%, FP=2%
  - Which one is better?
- IDS operation point
  - Point1: FN=1%, FP=2%
  - Point2: FN=10%, FP=0.5%
  - Which point to configure?

ROC Curve



Lesson: ROC curve provides tradeoff, but itself cannot tell you which one is better in many cases!

Cost-based Analysis

- Assign different costs to FP, FN according to the risk model in operation environment
- Compute the expected cost
    - the operation point/IDS with the minimal expected cost is better

Analysis on One Example

- [GU,Oakland'01] Using a decision tree model, the expected cost of operating at a given point on the ROC curve is the sum of the products of the probabilities of the IDS alerts and the expected costs conditional on the alerts

$$C_{exp} = Min\{C\beta B, (1-\alpha)(1-B)\} + Min\{C(1-\beta)B, \alpha(1-B)\}$$

- C is a cost ratio: C=Cost(FN)/Cost(FP)
- B is the base rate P(I)

Analysis on One Example (cont.)

$$C_{exp} = CB \qquad if\ CB < \frac{\alpha}{1-\beta}$$

Improvement of FP,FN does not show effect!

$$C_{exp} = C\beta B + \alpha \qquad if\ \frac{\alpha}{1-\beta} < CB < 1$$

Improvement of FN and change of base rate do not show effect!

$$C_{exp} = 1 + \alpha \qquad if\ CB > 1$$

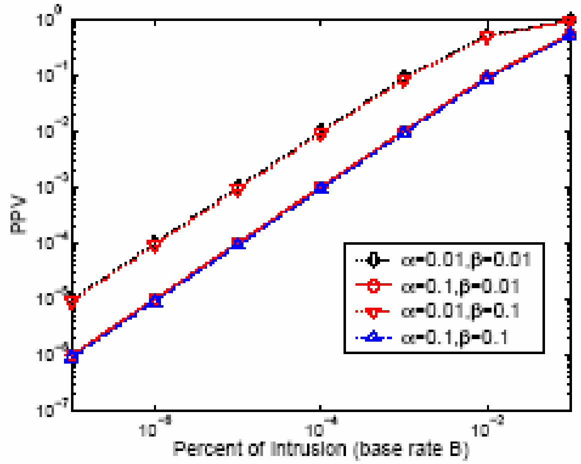
Problem with Cost-based Analysis

- Cost measures (cost of false alarms and missed attacks) determined *subjectively*, and usually they are very *hard* to choose accurately
  - Lack of good risk analysis models in many real situations
  - So it cannot be used to *objectively* evaluate and compare IDSs
- It does not provide an intrinsic measure of detection performance (or accuracy)

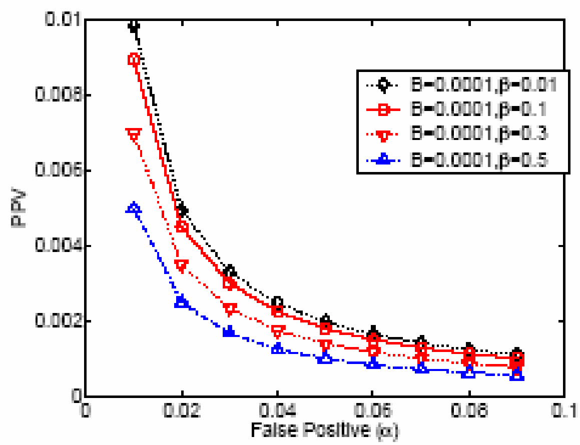
Two Other Metrics

- Consider the important environment parameter, base rate. And from a user point of view
- PPV: Positive predictive value, or “Bayesian detection rate”
  - P(I|A): given IDS alerts, how many of them are real intrusions?
- NPV: Negative predictive value
  - P(¬I|¬A): given there are no IDS alerts, does it mean there are really no intrusions?
- Base rate fallacy [Axelsson,CCS99]: PPV is very low because B is extremely low in realistic environment
- Tradeoff is also needed between PPV and NPV

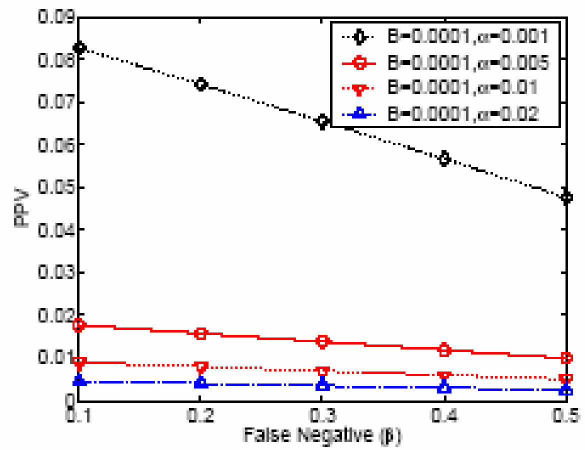
PPV, NPV



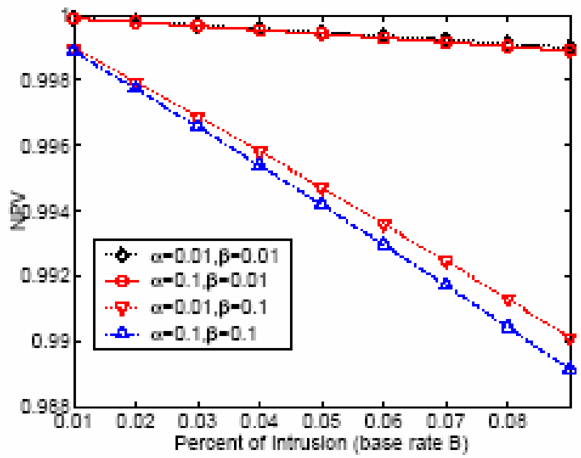
(a) PPV in Realistic Environment (both axes in log-scale)



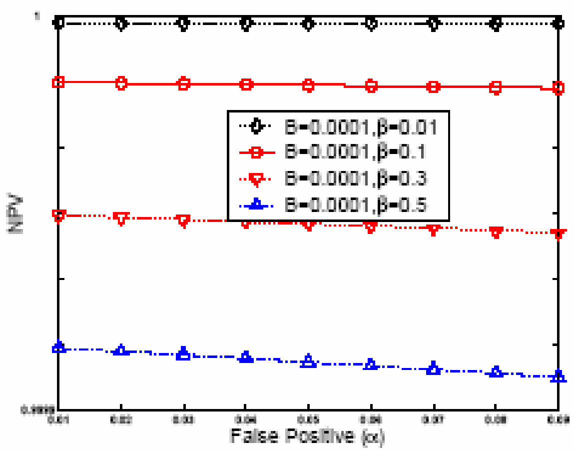
(b) The Effect of FP for PPV



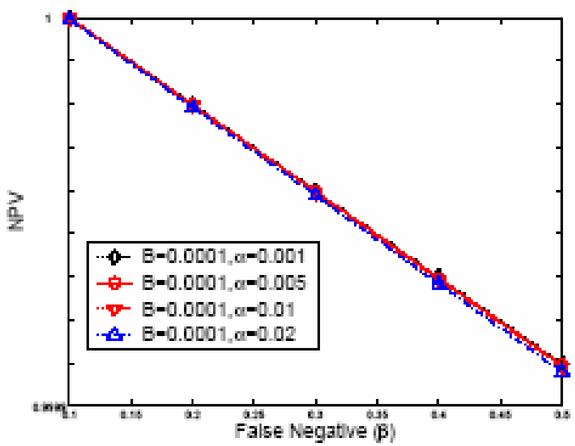
(c) The Effect of FN for PPV



(d) NPV in Realistic Environment



(e) The Effect of FP for NPV



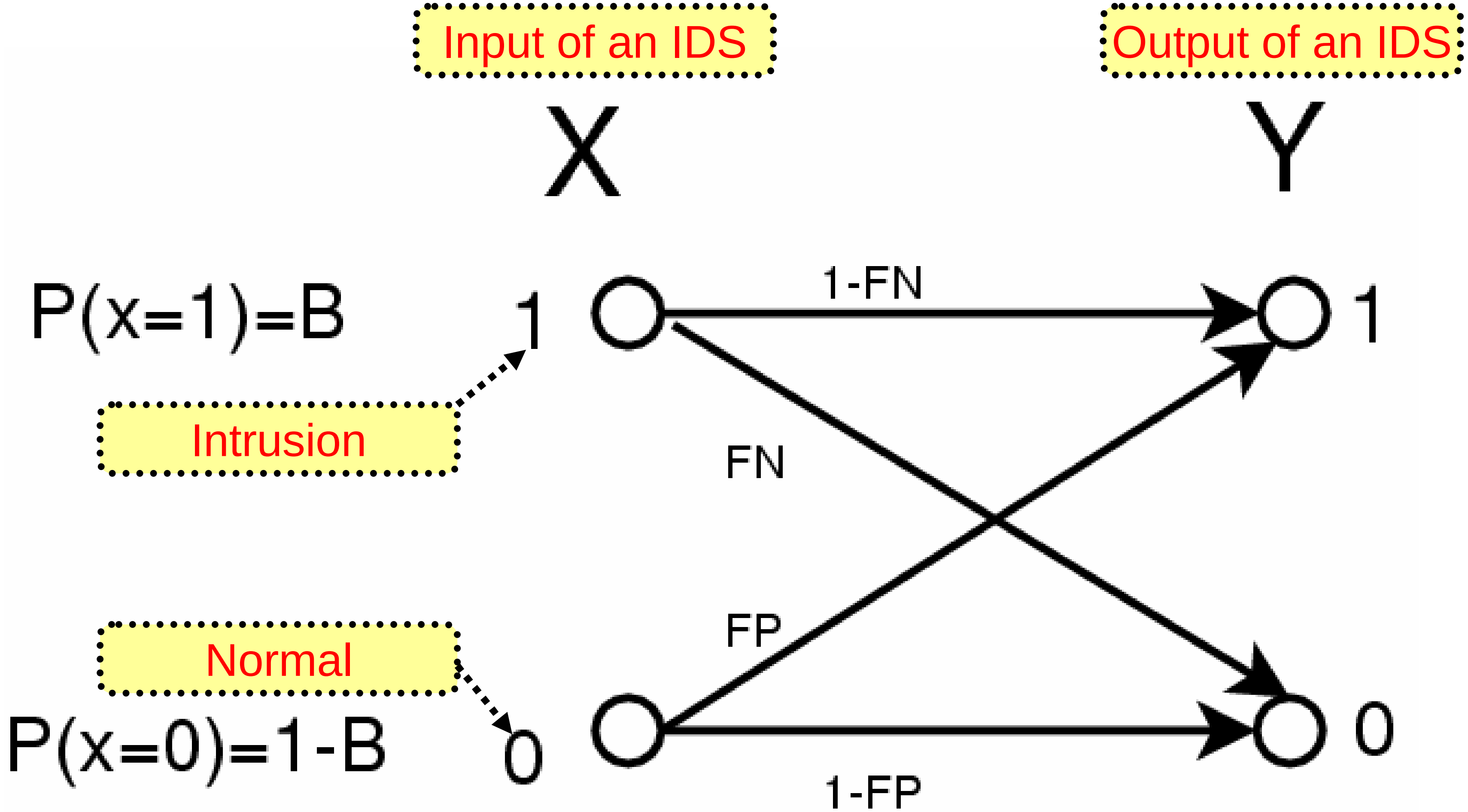
(f) The Effect of FN for NPV

What We Want?

- A single unified metric that takes into account all the important aspects of detection capability
- Be objective, not depend on any subjective measure (which is hard to determine in many realistic situations)
- Be sensitive to IDS operation parameters to facilitate fine tuning and fine-grained comparison of IDSs

C_{ID}

An Information-Theoretic View of Intrusion Detection



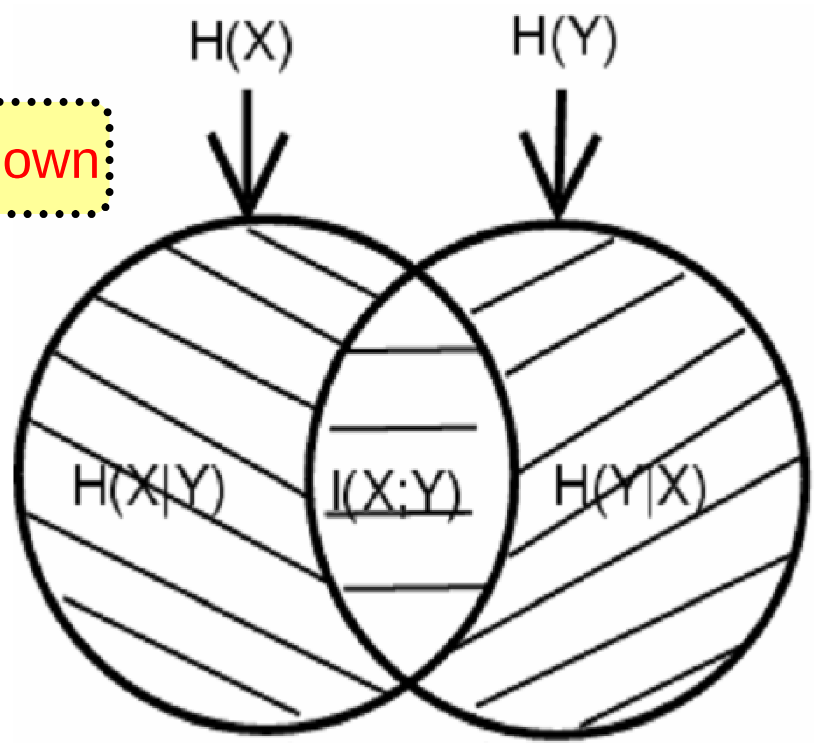
Information Theory Background

Uncertainty (information) of X

- Entropy H(X)

The remaining uncertainty in X after Y is known

- Conditional entropy H(X|Y)
- Mutual information I(X;Y)



The amount of reduction of uncertainty in X after Y is known

An Information-Theoretic View of Intrusion Detection (cont.)

- The purpose of an IDS (abstract level)
  - Classify the input correctly as normal or intrusion
  - The IDS output should faithfully reflect the ``truth'' about the input (whether there is an intrusion or not).

- Information-theoretic point of view, we should have less *uncertainty* about the input given the IDS output

- **Mutual information**: captures the reduction of original uncertainty (intrusion or normal) given that we observe the IDS alerts.

Intrusion Detection Capability

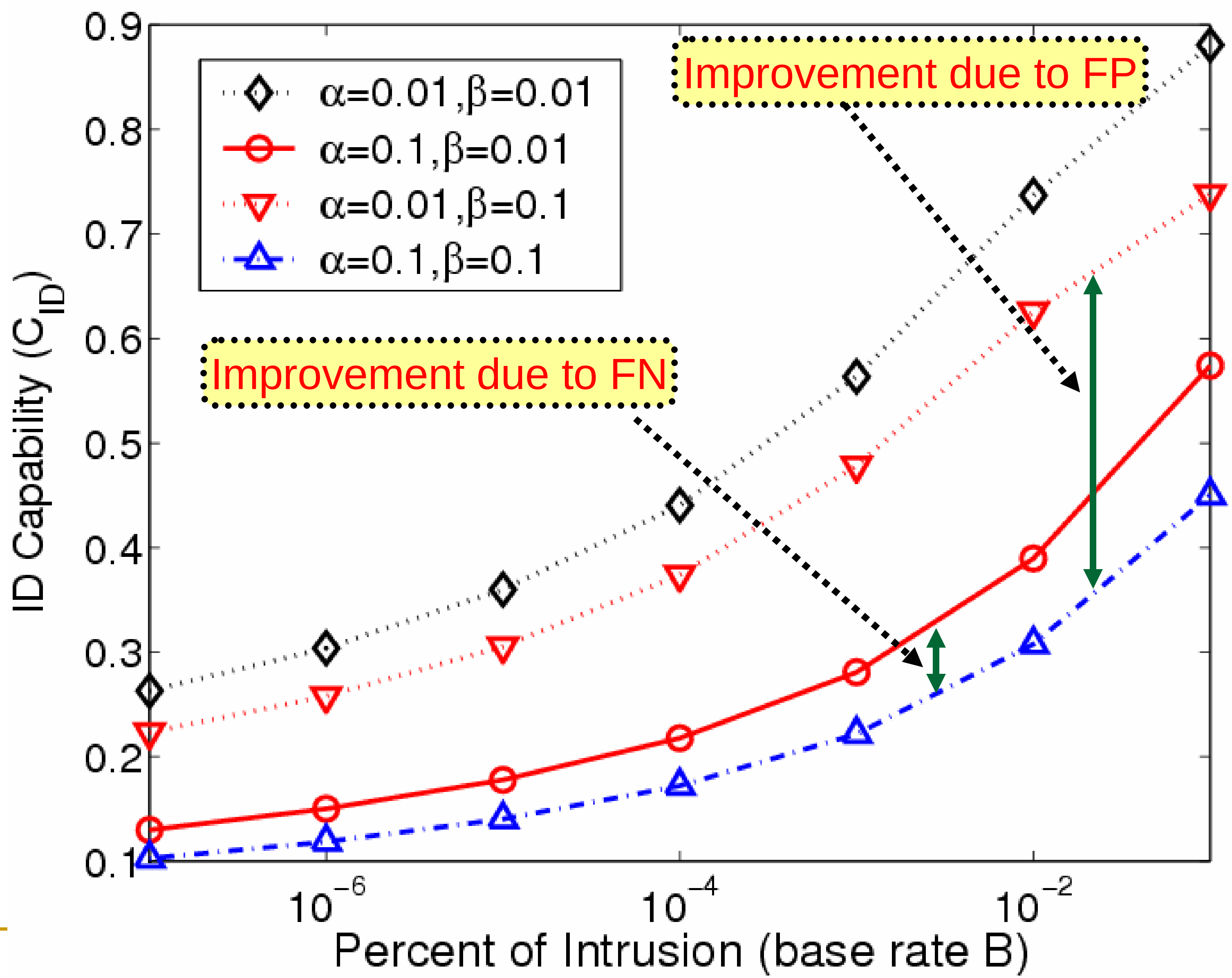
$$C_{ID} = \frac{I(X;Y)}{H(X)}$$

$$C_{ID} = I(X;Y)/H(X) = (H(X) - H(X|Y))/H(X)$$

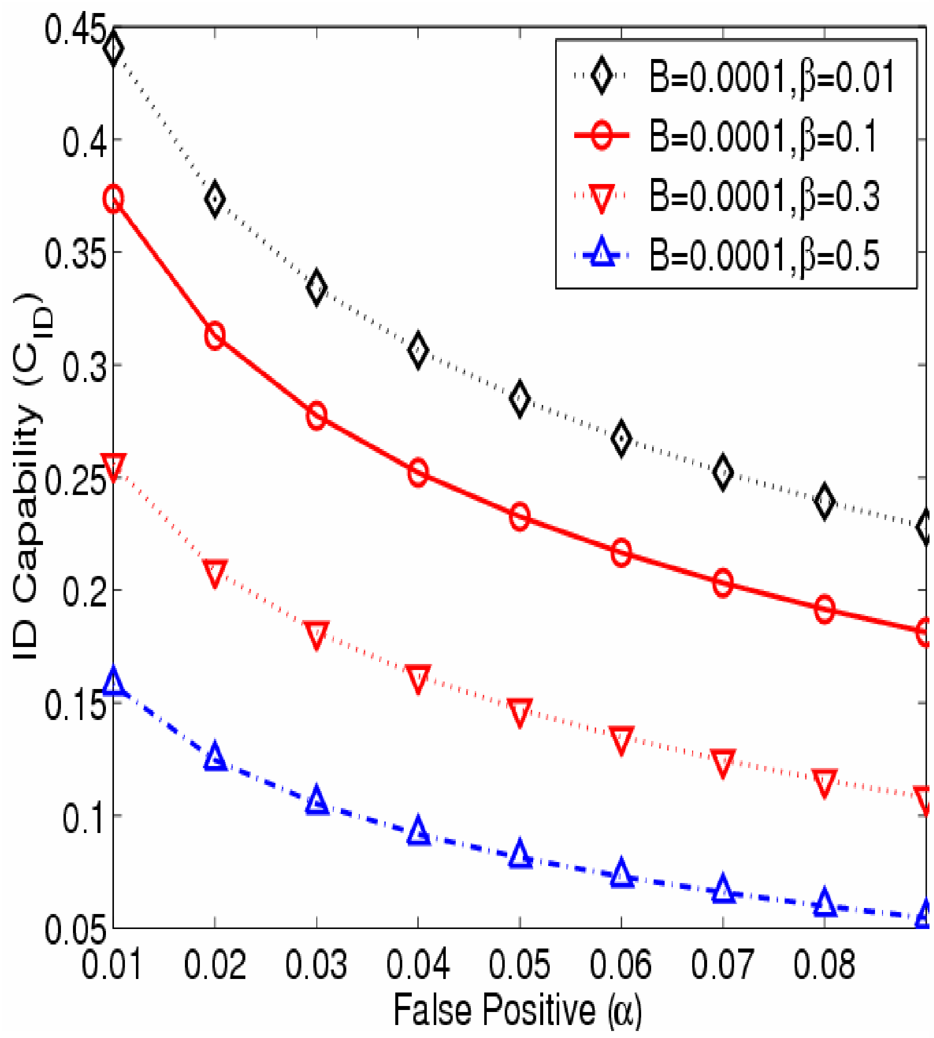
- A function of three basic variables
  - B
  - FP
  - FN

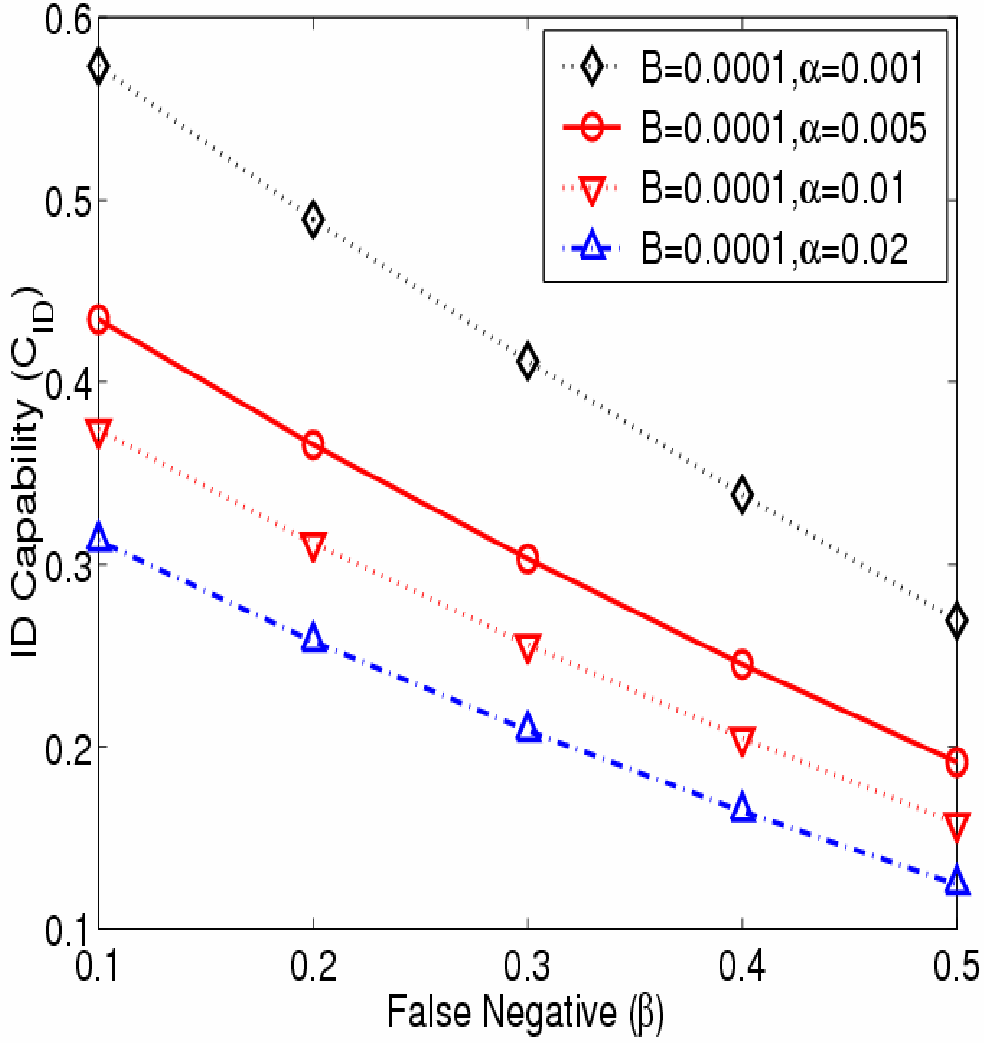
Another Intuitive Meaning

- Input X: a data stream (a stochastic binary vector with the ground truth indication unknown to the IDS)
- Output Y: an alert stream that should ideally be identical to X
- The IDS has to make correct guesses about the unknown X
- The actual number of required binary guesses is H(X) (the ``real'' information content of X). Of these, the number correctly guessed by the IDS is I(X;Y). (see Venn diagram for the intersection H(X) and H(Y))
- Thus I(X; Y)/H(X) is the fraction of correct 'information' guesses



0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
ID Capability (C_{ID})
10^{-6}
10^{-4}
10^{-2}
Percent of Intrusion (base rate B)
α=0.01,β=0.01
α=0.1,β=0.01
α=0.01,β=0.1
α=0.1,β=0.1
Improvement due to FP
Improvement due to FN


ID Capability (C_{ID})
False Positive (α)
B=0.0001,β=0.01
B=0.0001,β=0.1
B=0.0001,β=0.3
B=0.0001,β=0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08


ID Capability (C_{ID})
False Negative (β)
B=0.0001,α=0.001
B=0.0001,α=0.005
B=0.0001,α=0.01
B=0.0001,α=0.02
0.6
0.5
0.4
0.3
0.2
0.1
0.1
0.2
0.3
0.4
0.5

Other Similar Metrics

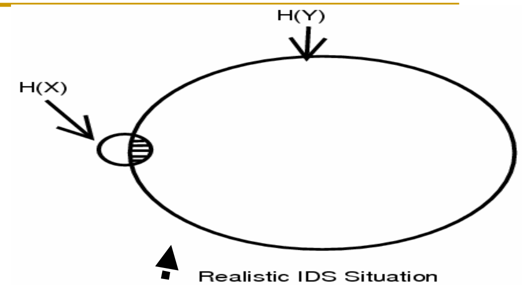
- Based on different ways to normalize mutual information

$$NMI = (H(X) + H(Y))/H(X,Y)$$

$$NMI' = I(X;Y)/H(X,Y)$$

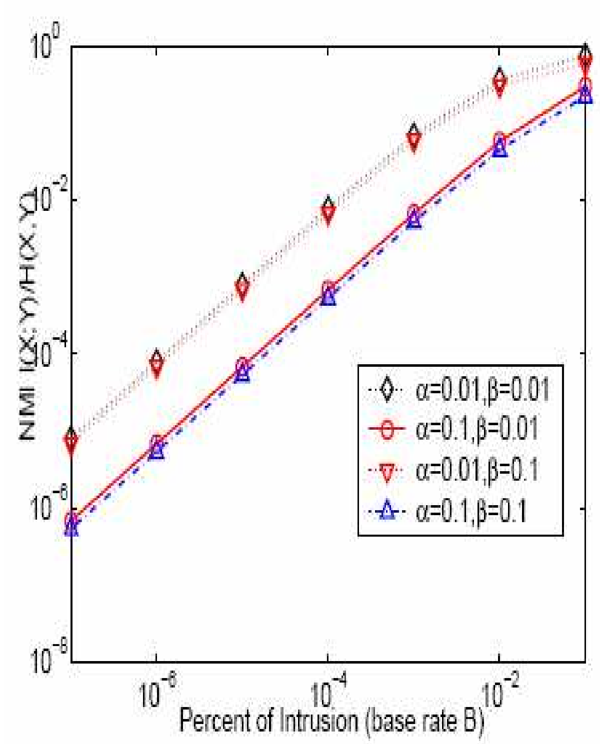
$$NAMI = I(X;Y)/H(Y)$$

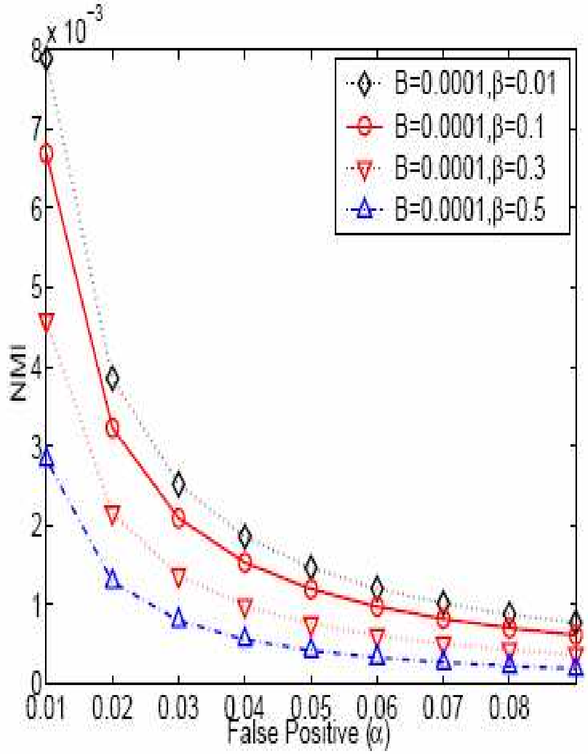
NMI' as an Example

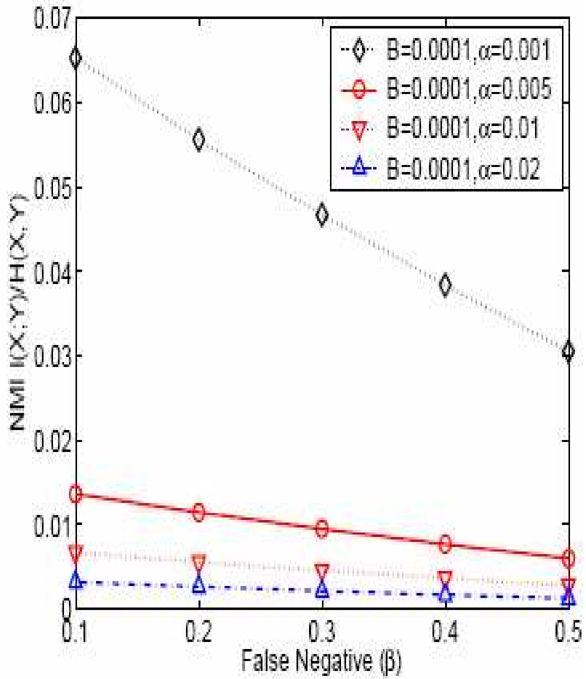


- Less sensitive (note the *orders of magnitude* difference in scales with C_{ID})

Reason







Why C_{ID} is better than existing metrics?

- A clear information-theoretic meaning
  - Not arbitrary subjective cost setting
- A unified metric
  - A nature tradeoff by taking care of existing metrics
- A more sensitive metric
  - Good to demonstrate the effect of the subtle changes of intrusion detection systems

Unified Metric: C_{ID}

$$C_{ID} = I(X;Y)/H(X) = (H(X) - H(X|Y))/H(X)$$

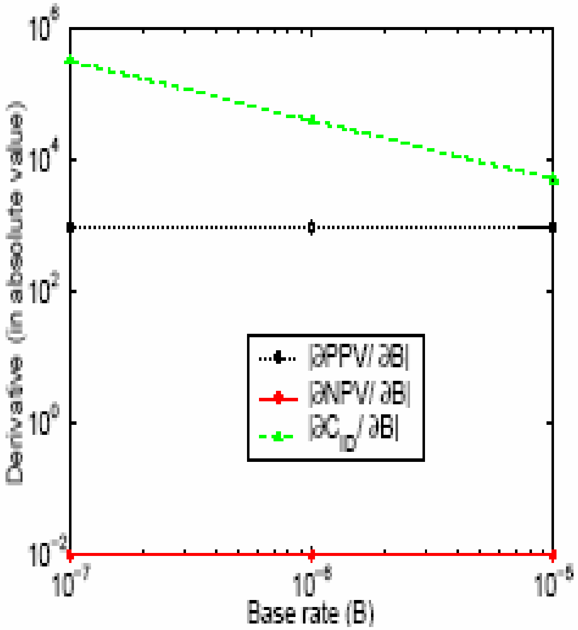
$$H(X) = -\sum_{x} p(x)\log p(x) = -B\log B - (1-B)\log(1-B)$$

$$H(X|Y) = -B(1-\beta)\log PPV - B\beta\log(1-NPV) - (1-B)(1-\alpha)\log NPV - (1-B)\alpha\log(1-PPV)$$

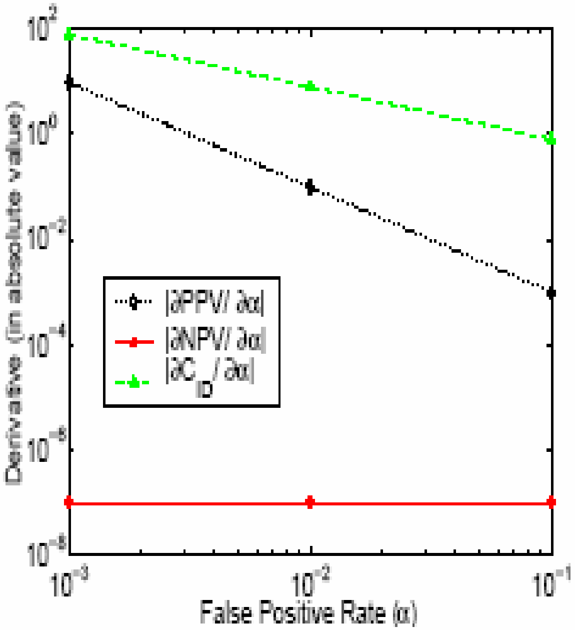
Unify existing metrics;
Also can be viewed as a nature cost tradeoff with the log() as cost functions

Sensitivity Analysis Using Derivatives

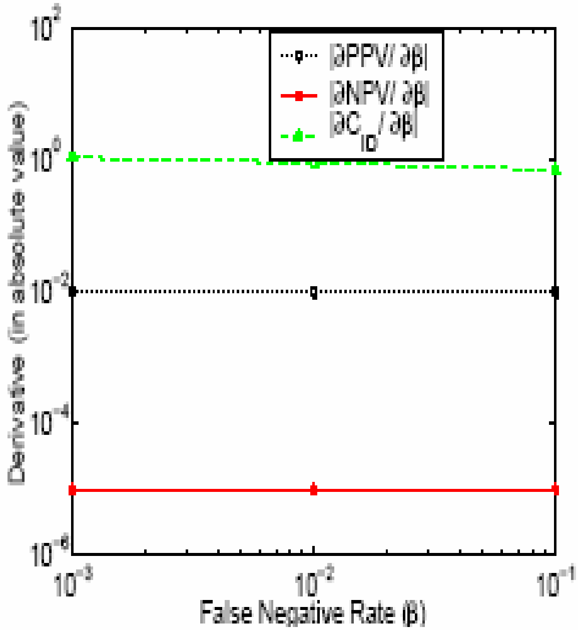
C_{ID} has the highest sensitivity compared to PPV, NPV



(a) Dependence on base rate analysis ($\alpha = 0.001, \beta = 0.01$)

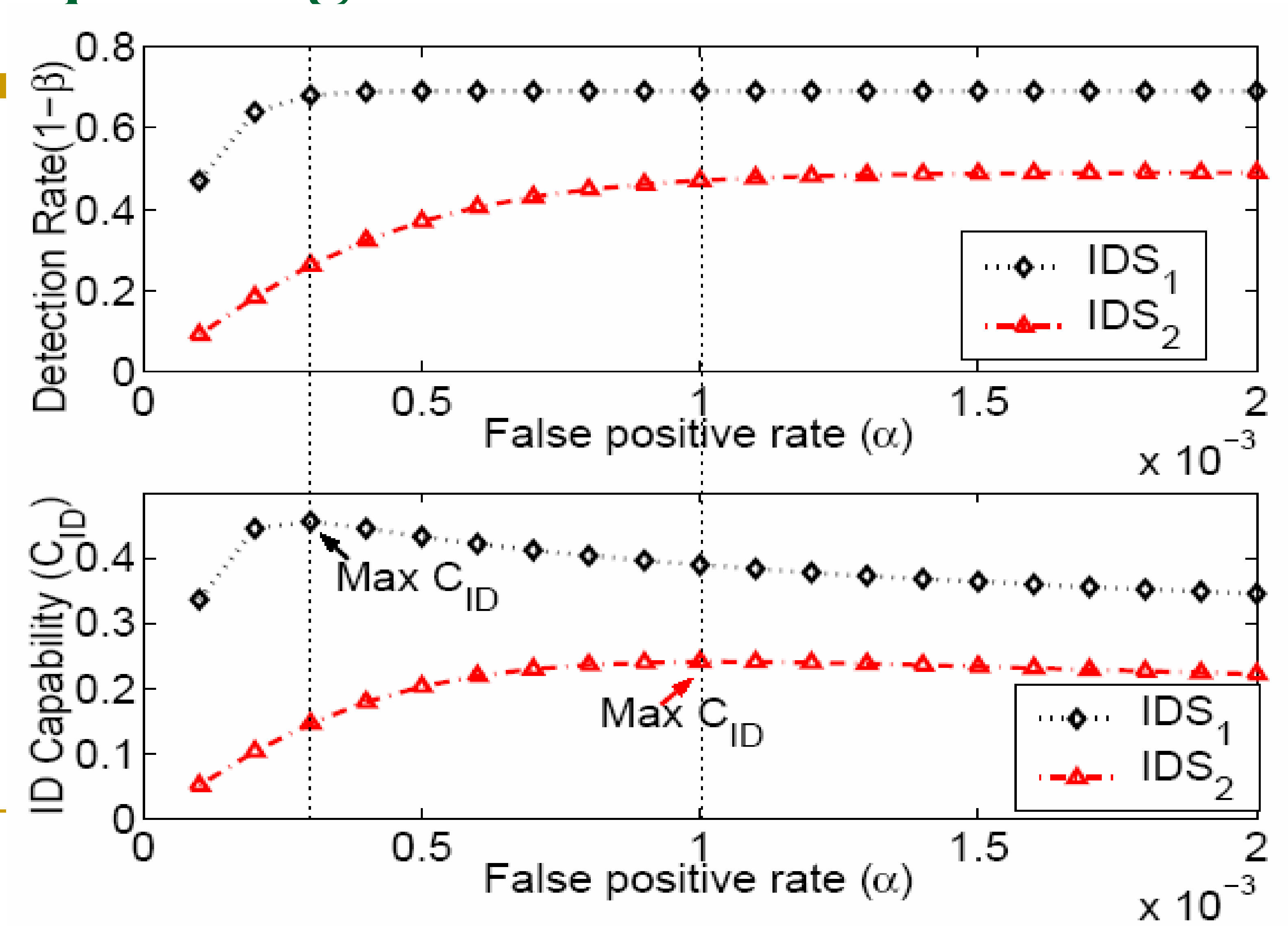


(b) Dependence on false positive rate analysis ($B = 0.00001, \beta = 0.01$)



(c) Dependence on false negative rate analysis ($B = 0.00001, \alpha = 0.001$)

Utility of C_{ID}: Selection of Optimal Operating Point



Utility of C_{ID}: Comparison of Different IDSs

- Example
  - IDS1: FP=1/660,000, TP=0.88
  - IDS2: FP=7/660,000, TP=0.97
- Which one is better?

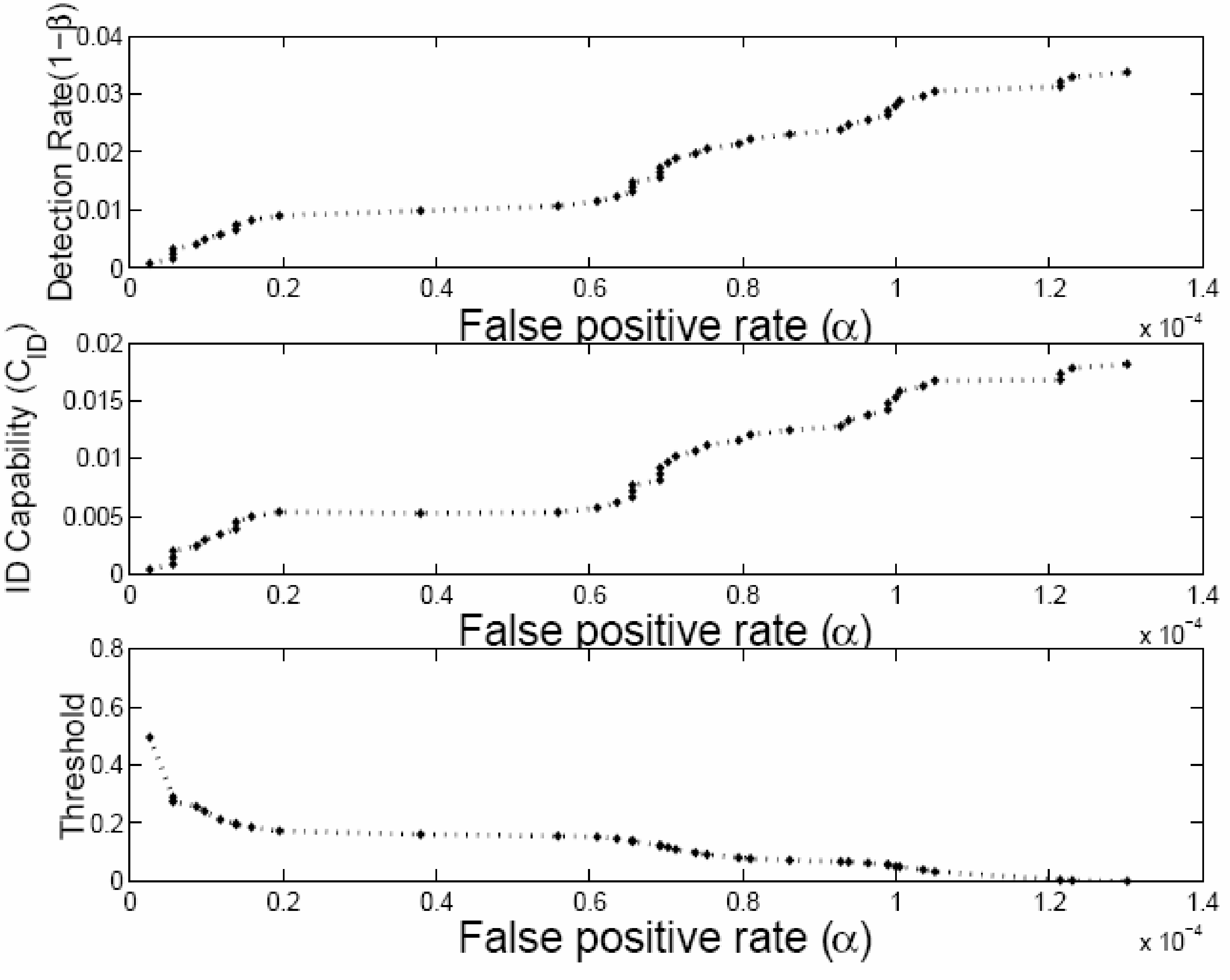
C_{ID} =0.8390

C_{ID} =0.8881

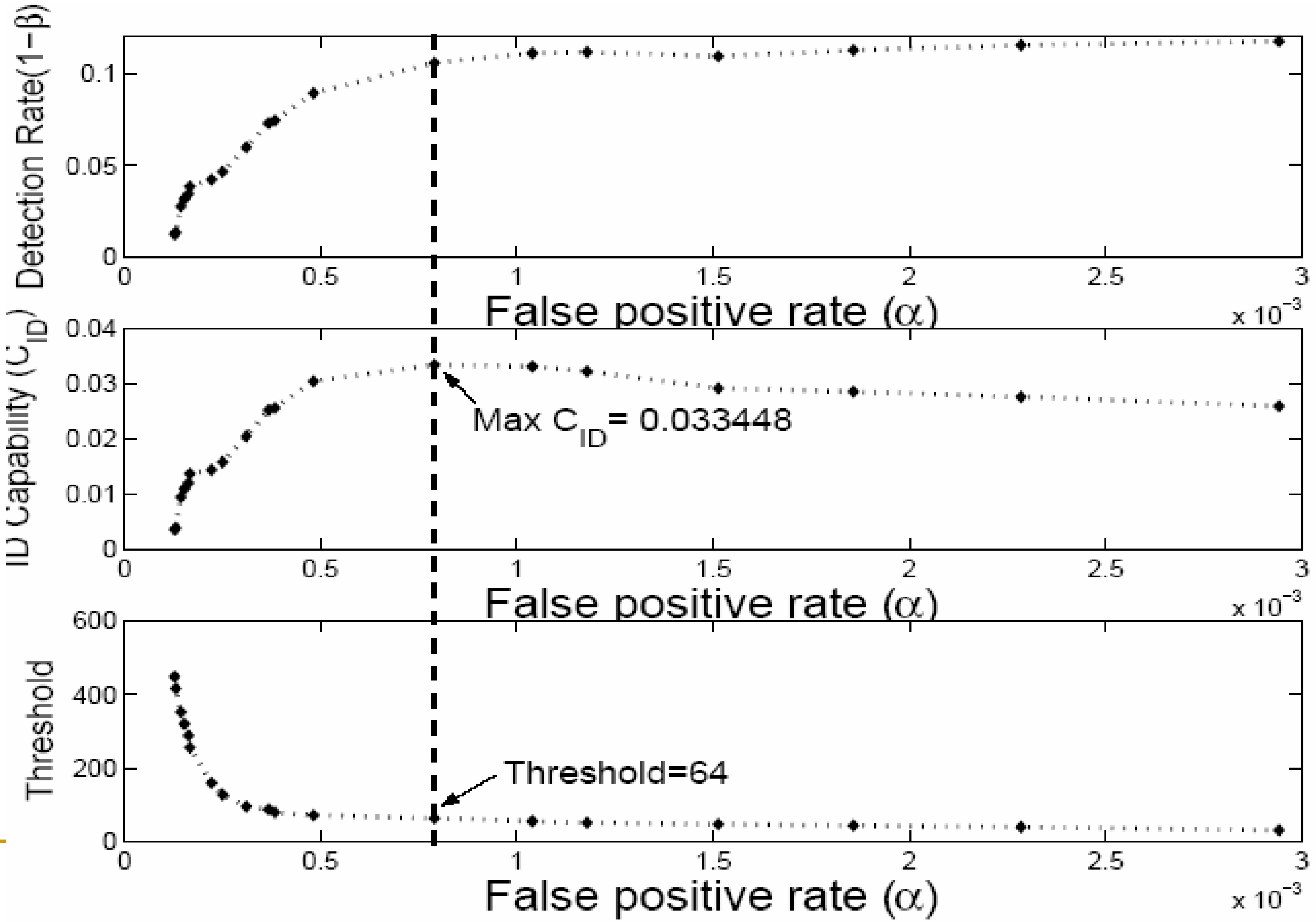
Real IDS Experiment

- Data
  - DARPA 1999 intrusion detection test data set
  - Georgia Tech CoC http traffic (about 6 hours)
- IDS
  - PHAD: Packet Header Anomaly Detection
  - PAYL: Payload Anomaly Detection
  - Snort (Version 2.1.0 Build 9)

PHAD


Detection Rate(1−β)
0.04
0.03
0.02
0.01
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
False positive rate (α)
x 10^-4
ID Capability (C_ID)
0.02
0.015
0.01
0.005
0
False positive rate (α)
x 10^-4
Threshold
0.8
0.6
0.4
0.2
0
False positive rate (α)
x 10^-4

PAYL


Detection Rate(1−β)
0.1
0.05
0
0
0.5
1
1.5
2
2.5
3
False positive rate (α)
x 10^-3
ID Capability (C_ID)
0.04
0.03
0.02
0.01
Max C_ID= 0.033448
False positive rate (α)
Threshold
600
400
200
Threshold=64
False positive rate (α)

Comparison

- PAYL
  - optimal operating threshold of 64
  - C_{ID}=0.033448
  - $\alpha=0.7\times10^{-3}$, $1-\beta=0.10563$
- Snort
  - $\alpha=0.0000006701$
  - $1-\beta=0.0117$
  - C_{ID}=0.0081

Better compared to PAYL

worse

worse

Summary and Future Work

- In-depth analysis of existing IDS metrics
- Studied the intrusion detection from the viewpoint of information theory
  - Proposed a *novel, natural, unified, objective, sensitive* metric to measure the capability of IDS
- Impact
  - Choose the best (optimized) operation point of an IDS
  - Compare different IDSs
- Future work
  - Rich encoding of X and Y
  - Analyze and improve both *internal* and *external* designs of IDS, by looking into multiple (chained) channel/layer architecture of the IDS

Q &A

Thank you!

Other Issues

- Estimation of Base Rate, FP, FN
- Unit of Analysis
- Involving Cost Analysis in C_{ID}