Principled Reasoning and Practical Applications of Alert Fusion in Intrusion Detection Systems

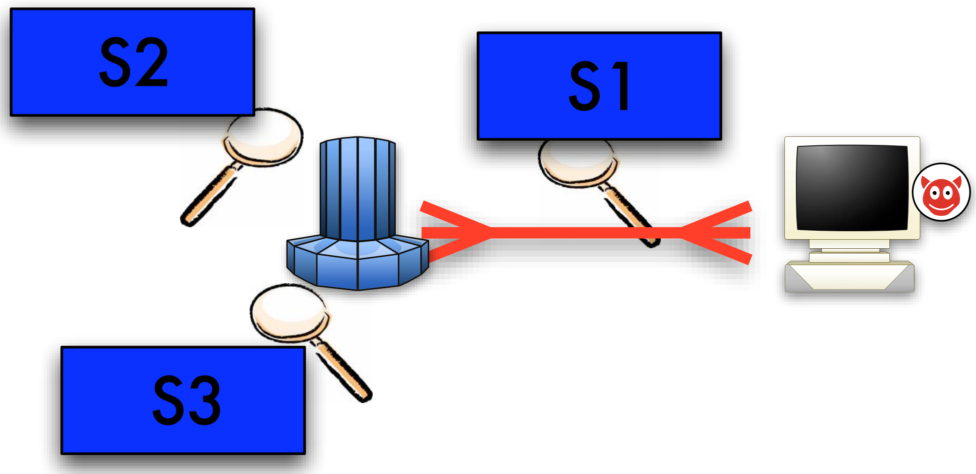
Guofei Gu, Alvaro A. Cárdenas, Wenke Lee

Georgia Institute of Technology

University of California, Berkeley

"OR" Rule for Combining Alerts

- Alerts of the same event can be raised by different methods
  - Input string length
  - Character distribution
  - Token finder etc...



- OR Rule:
  - Alert iff S1 OR S2 OR S3 Alerts
- Analyst is overwhelmed by the number of alarms
- String length might give many false alerts

"AND" Rule for Combining Alerts

- Polygraph[1]: Automatic Generation o[f] Signatures.
- Signature: Conjunction of all tokens
  - AND Rule:
    - Alert iff token1 AND token2 AND ... AND tokenN found in network flow.
- More false negatives: token observed i[n] suspicious, but not in every real worm

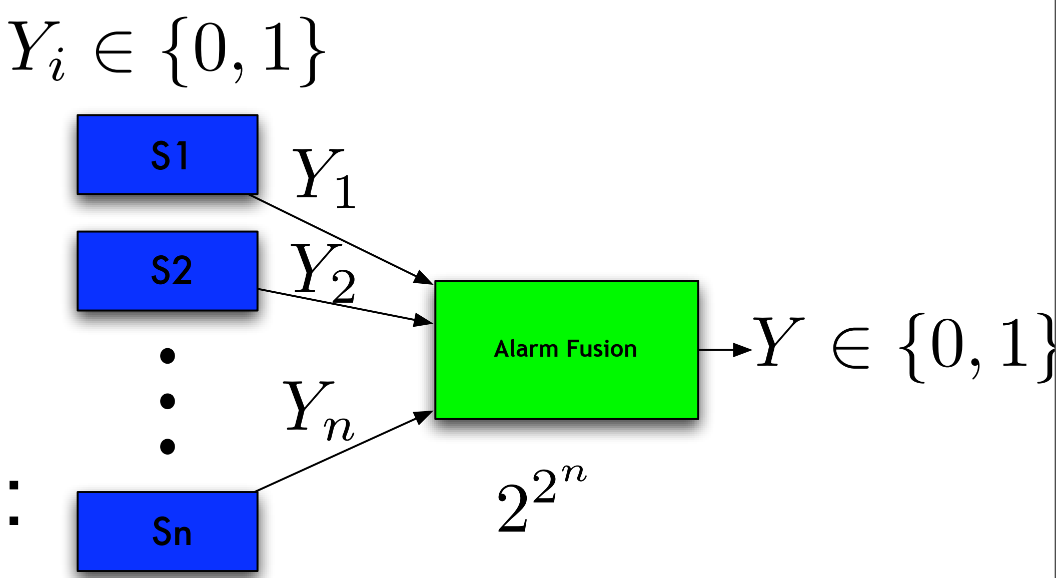
Tokens: All distinct substrings of a minimum length:
e.g., If there are K occurrences of "http" "ttp" will not be considered unless it appears another K times and not as part of http

Token observed in all samples of the suspicious flow, but does not appear in every sample of the worm.

[1]. Newsome, Karp, Song. *Polygraph: Automatically Generating Signatures for Polymorphic Worms*. IEEE S&P, 2005

Our Goal: Study Design Space for Combining Alerts

- With n tokens (or sensors) there are 2^{2^n} possible fusion rules
- AND-rules and OR-rules are only 2 of them
- But there are many more: Majority voting, Select only one, etc...



Which Fusion Rule is the Best?

- We want to find the “best” fusion rule(s):

$$g^* = \arg \max_{f \in \{g:\{0,1\}^n \rightarrow \{0,1\}\}} \Phi(f)$$

- Problem 1: Find the rules that give an optimal ROC curve
- Problem 2: Find the rules that minimize the operational “cost” of an IDS
- Problem 3: Prioritize alerts

Our Solution: Likelihood Ratio Test (LRT)

- Each rule has a different False Alarm vs. False Negative tradeoff (we obtain a LRT estimate).
- LRT-Rule is optimal for Problem 1 (best ROC), Problem 2 (minimize costs) and Problem 3 (ranking of alarms).
- Principled (theoretically sound) and practical (useful and intuitive) way of combining intrusion detection sensors.

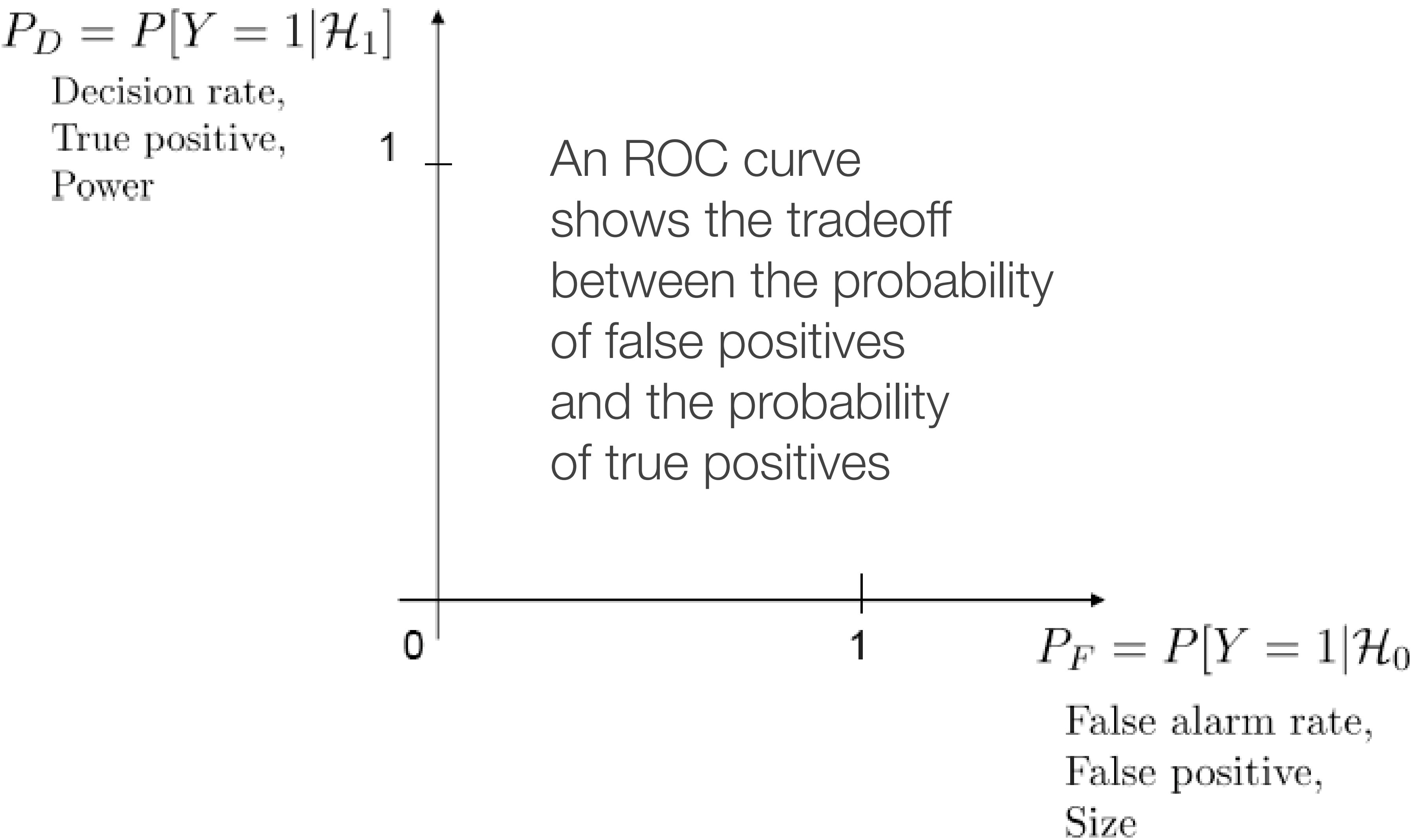
Agenda

- Metric 1: Optimal ROC curve
- Metrics 2 & 3: Minimum cost and ranking
- Experiments
- Conclusions and Future Work

Notation and Definitions

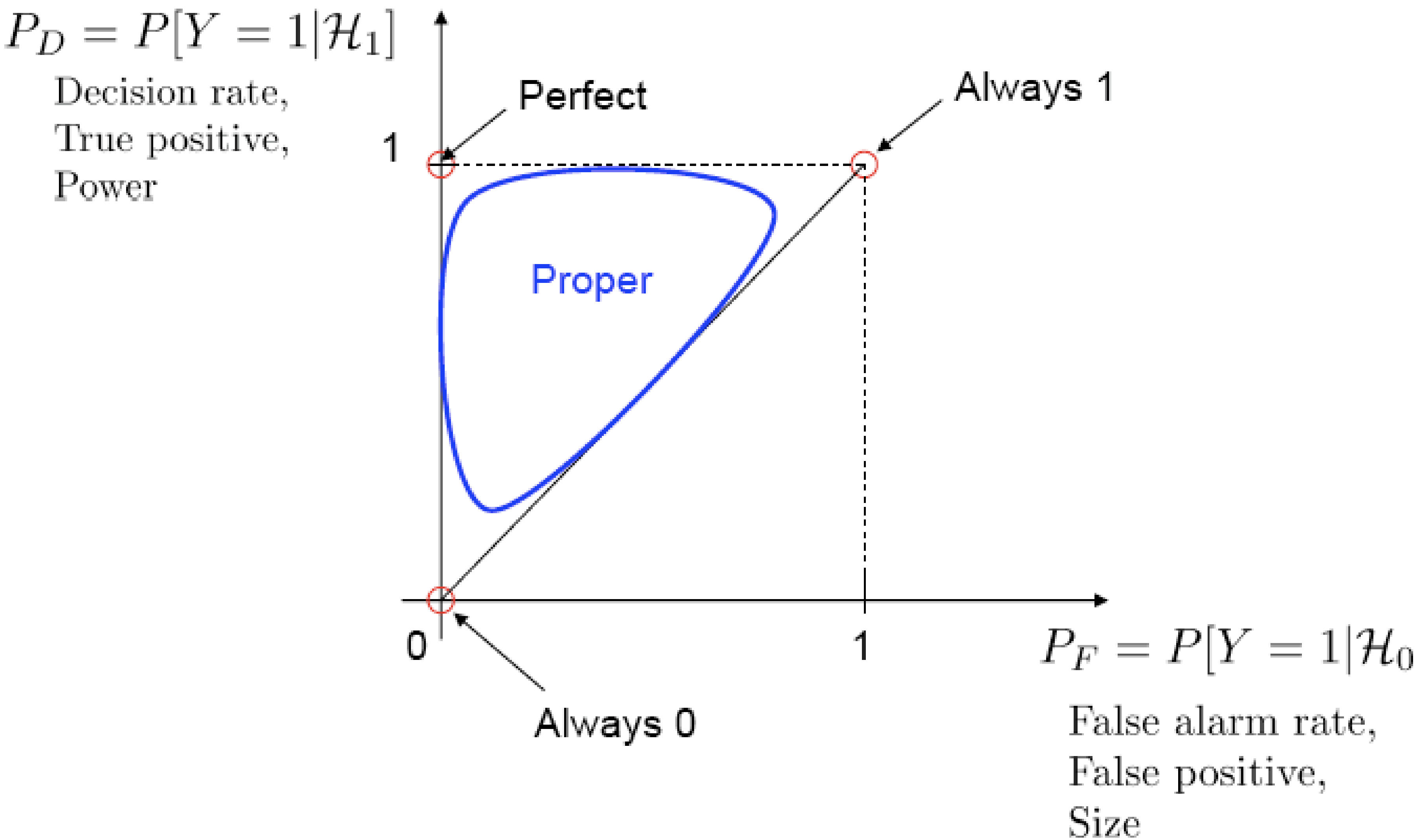
- Intrusion I=1, otherwise I=0
- Output is Y=1 (alarm), Y=0 (no alarm)
- P_F=Pr[Y=1|I=0] and P_D=[Y=1|I=1]
- There is a tradeoff between P_F and P_D
- The ROC curve shows points (P_{FA},P_D) for different "configurations" of an IDS

Metric 1: Receiver Operating Characteristic (ROC) Curve

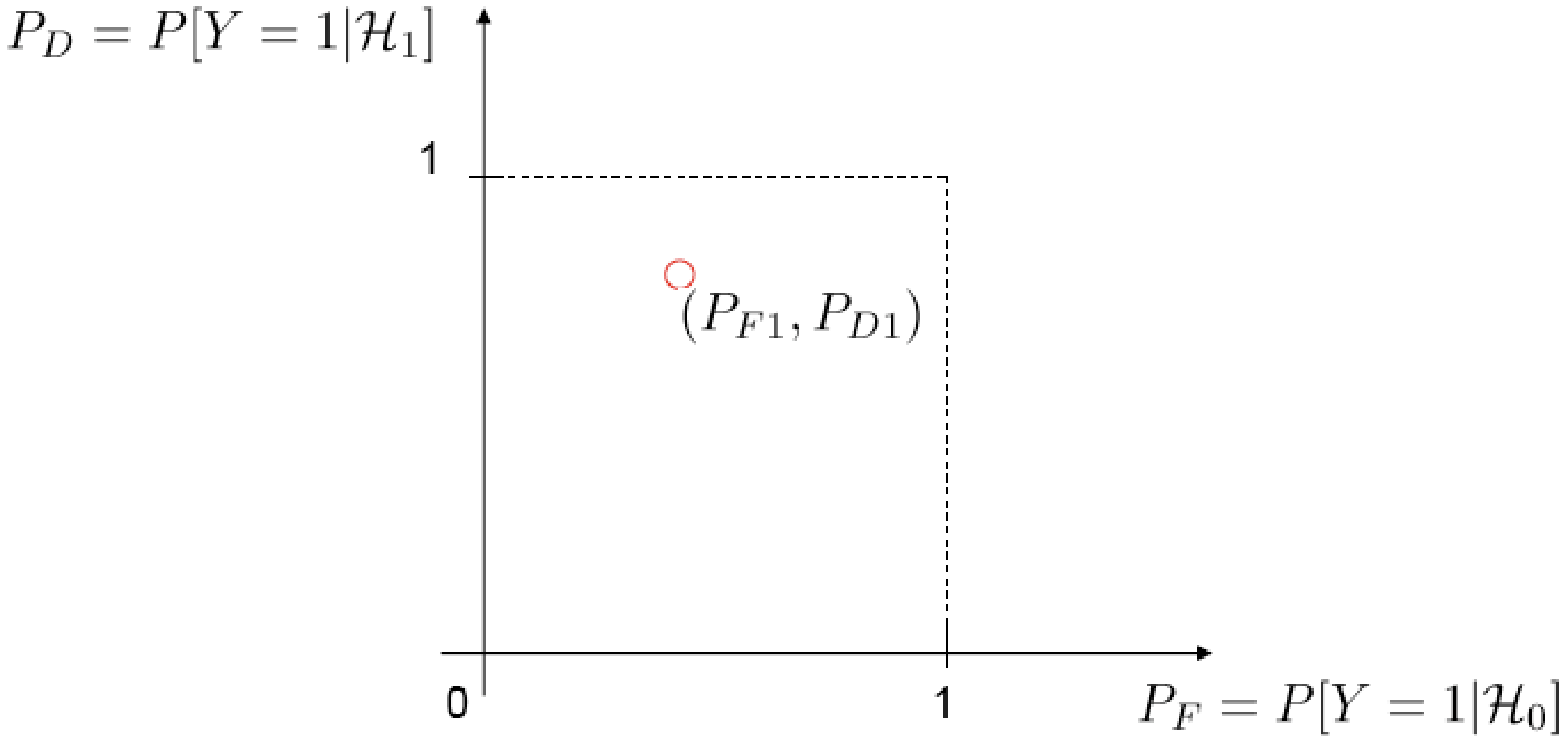


An ROC curve shows the tradeoff between the probability of false positives and the probability of true positives

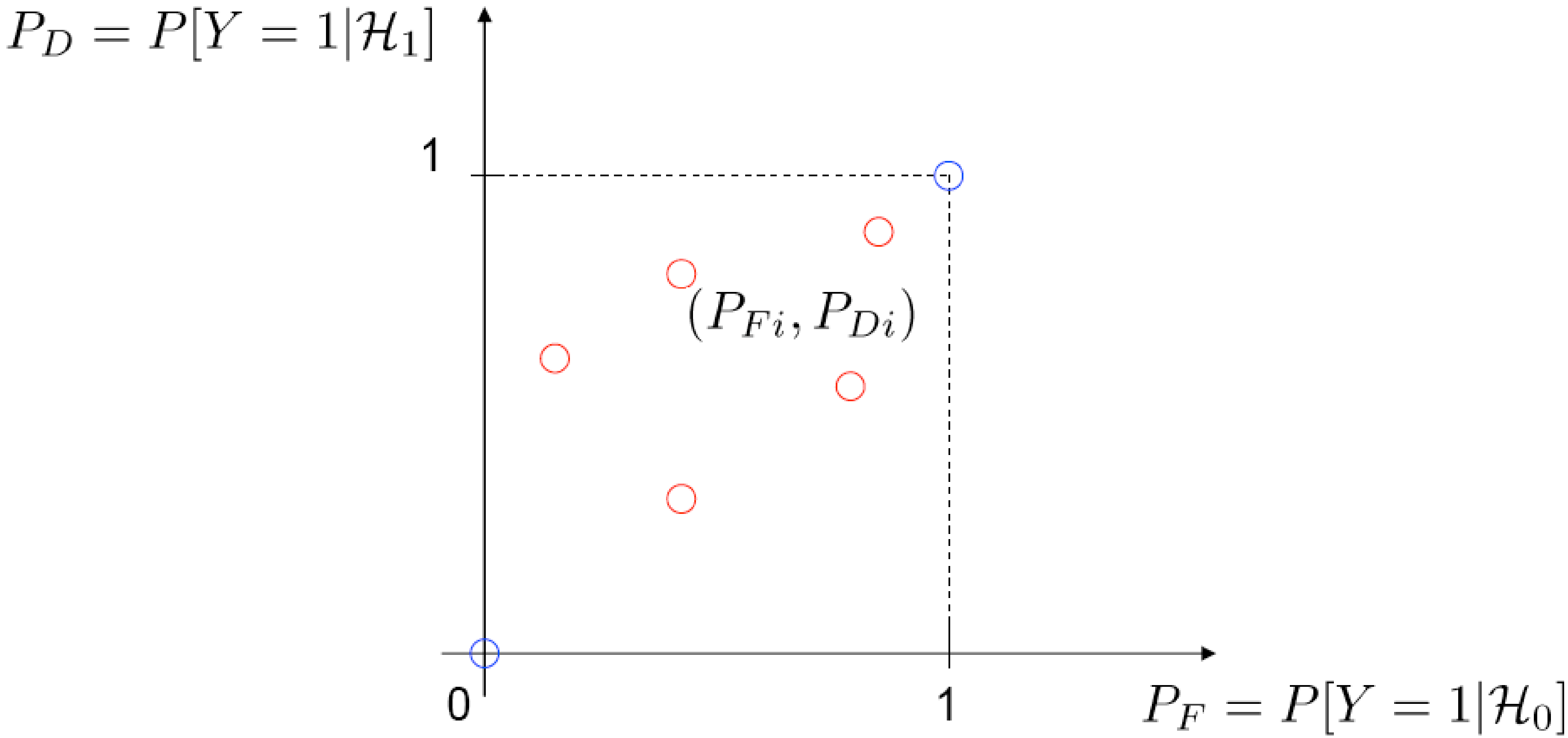
Metric 1: Receiver Operating Characteristic (ROC) Curve



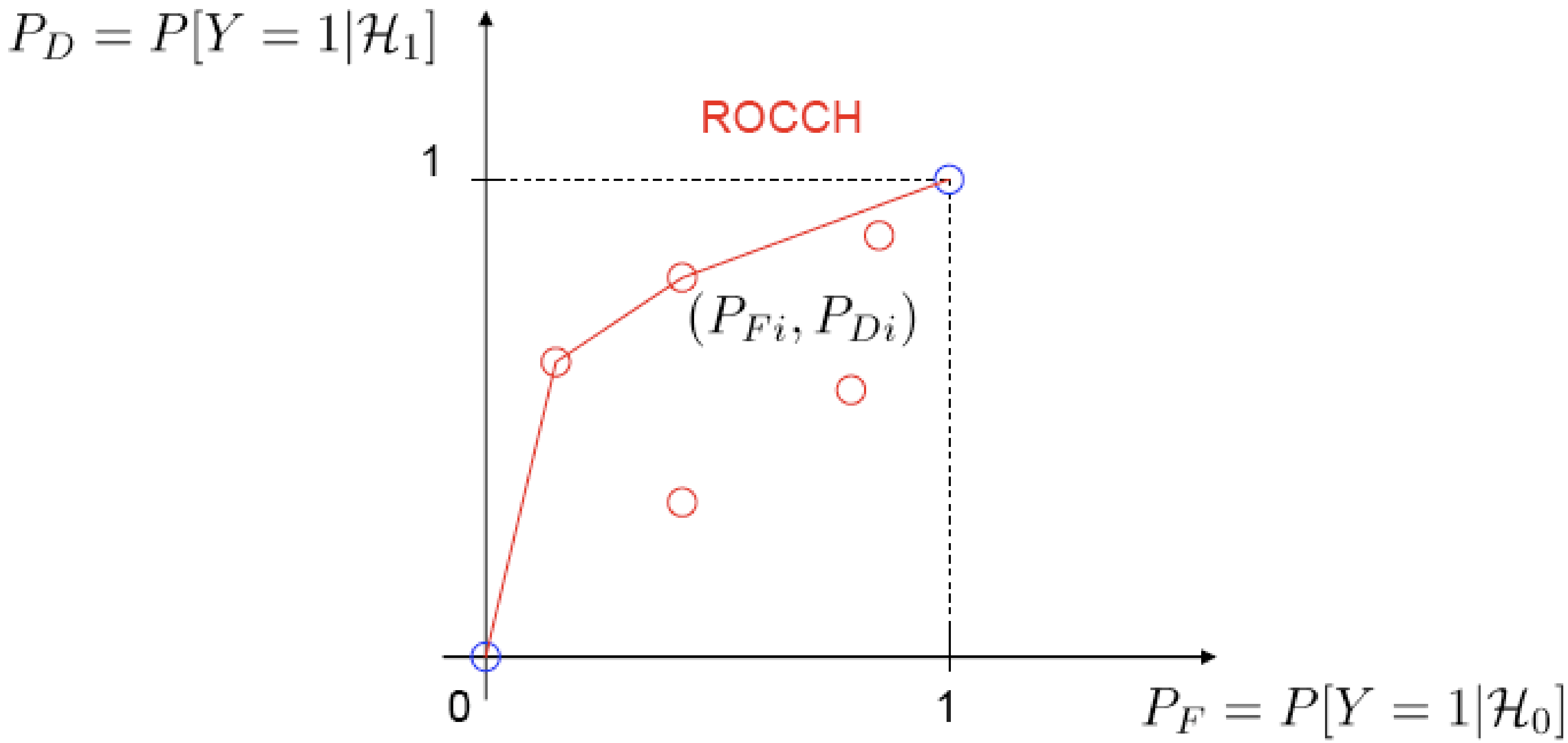
Performance of Sensor_1



P_{Fi} and P_{Di} estimates for multiple sensors

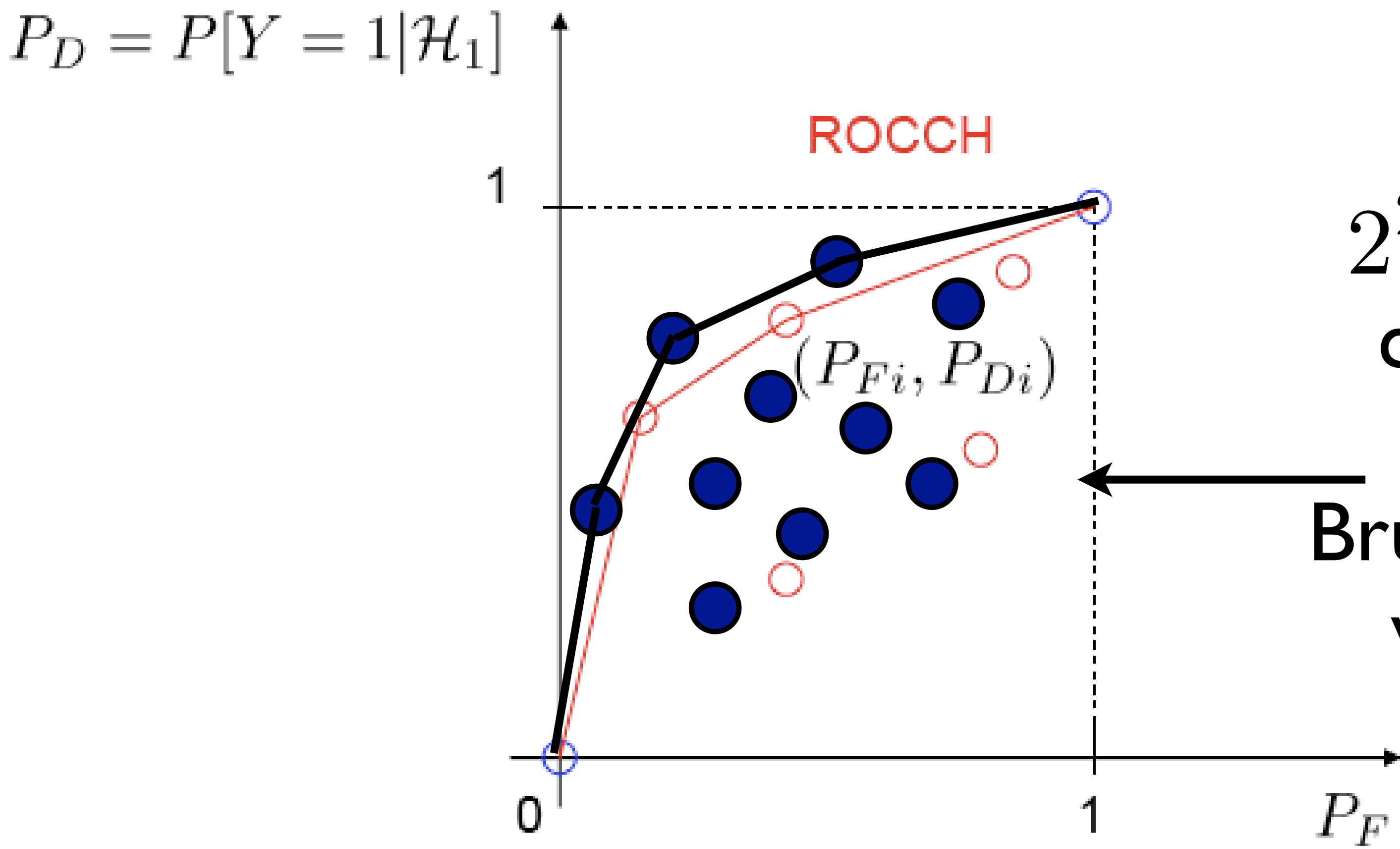


Previous Work: The ROC Convex Hull (ROCCH)[2]



[2]. Provost, Fawcett. Robust Classification for Imprecise Environments. Machine Learning 2001

ROCCH Gives Suboptimal ROC



2^{2^n} Possible combination rules

Brute for search won't work

Neyman-Pearson Theory

- Given observation Y: test Null Hypothesis H_0 vs. alternative H_1
- If we know $P(Y|H_0)$ and $P(Y|H_1)$, then the test $D(Y)$ that maximizes $P[D(Y)=H_1|H_1]$ for a fixed $P[D(Y)=H_1|H_0]$ is:

$$\mathcal{D}(\mathbf{Y}) = \begin{cases} 1 & \text{if } \ell(\mathbf{Y}) > \tau \\ \gamma & \text{if } \ell(\mathbf{Y}) = \tau \\ 0 & \text{if } \ell(\mathbf{Y}) < \tau \end{cases} ,$$

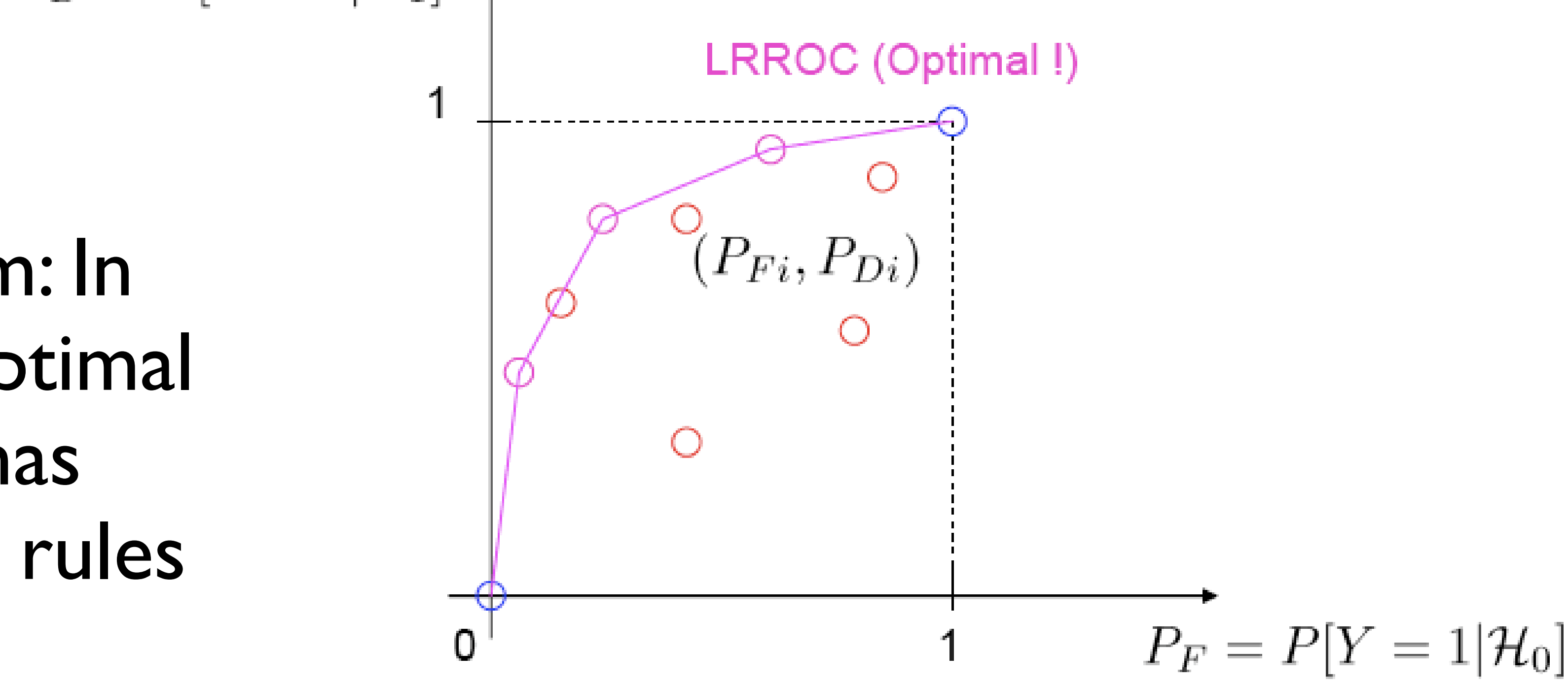
- Where $l(Y)= P(Y|H_1)/P(Y|H_0)$ is the likelihood ratio.

Our Work: The Likelihood Ratio Test for Fusing Alarms

$$\mathcal{D}(\mathbf{Y}) = \begin{cases} 1 & \text{if } \ell(\mathbf{Y}) > \tau \\ \gamma & \text{if } \ell(\mathbf{Y}) = \tau \\ 0 & \text{if } \ell(\mathbf{Y}) < \tau \end{cases},$$

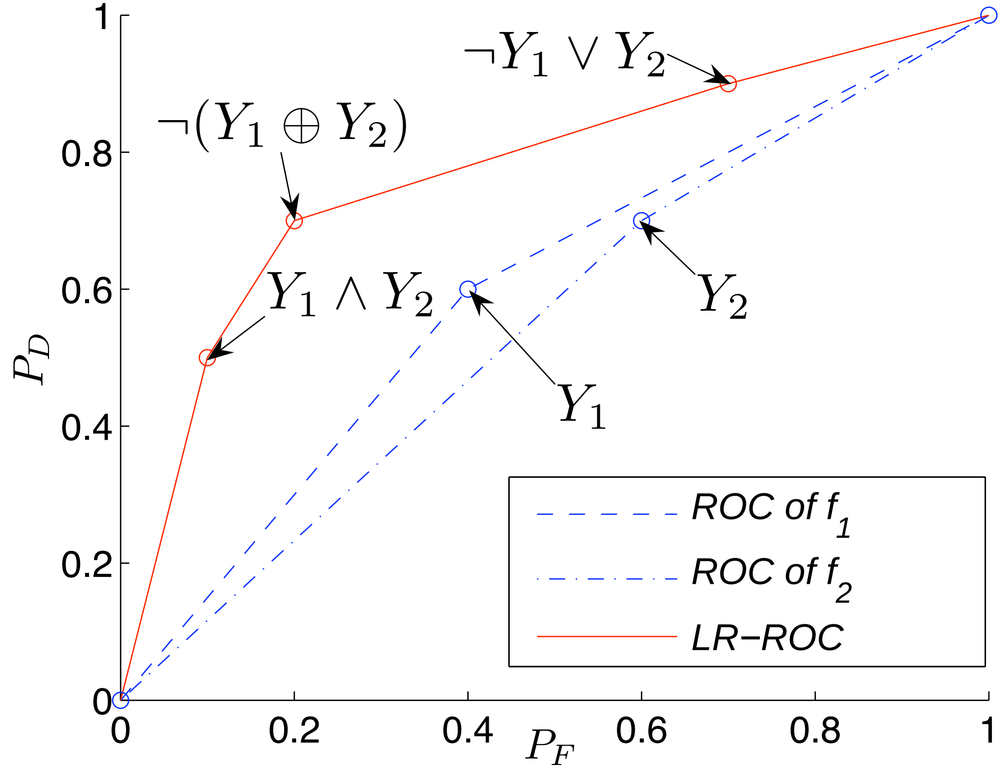
independence assumption $\ell(\vec{1}) = \dfrac{P_{D1} \ldots P_{Dn}}{P_{F_1} \ldots P_{Fn}}$

no independence $\ell(\vec{Y}) = \dfrac{\Pr[Y_1, Y_2, \ldots, Y_n | H_1]}{\Pr[Y_1, Y_2, \ldots, Y_n | H_1]}$



Theorem: In general, optimal ROC has $2^n + 1$ rules

Example of the Likelihood-Ratio Test



Class 1 (H_1)

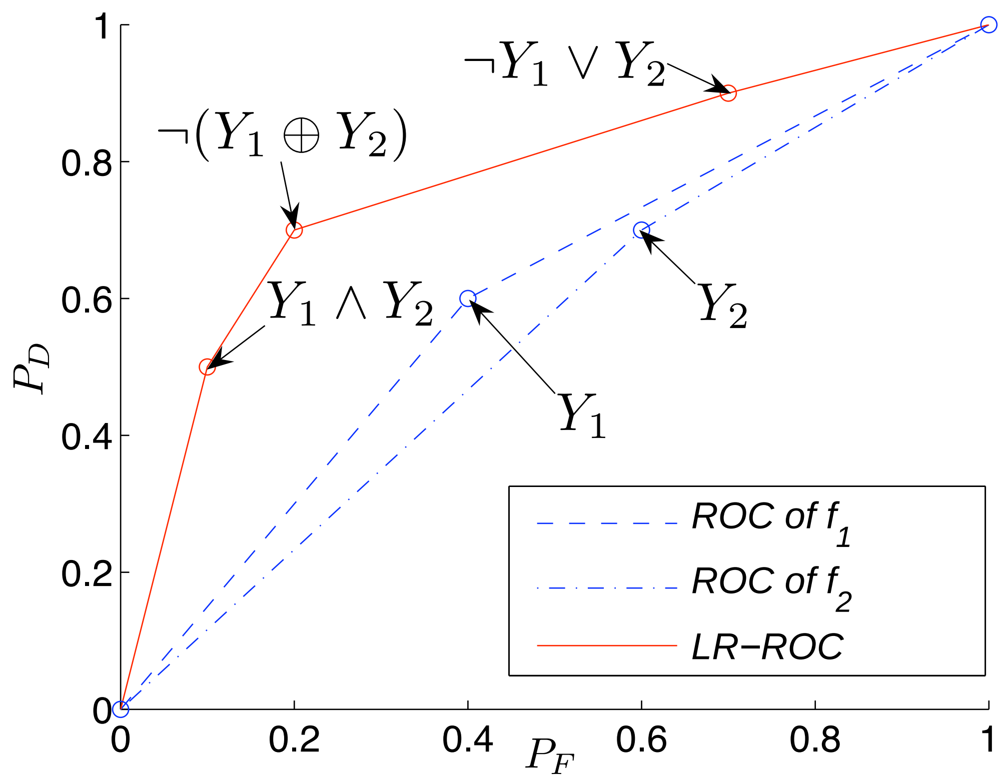
Y_2 \ Y_1	0	1
0	0.2	0.1
1	0.2	0.5

Class 0 (H_0)

Y_2 \ Y_1	0	1
0	0.1	0.3
1	0.5	0.1

$\ell(00)=2$

Example of the Likelihood-Ratio Test



Class 1 (H_1)

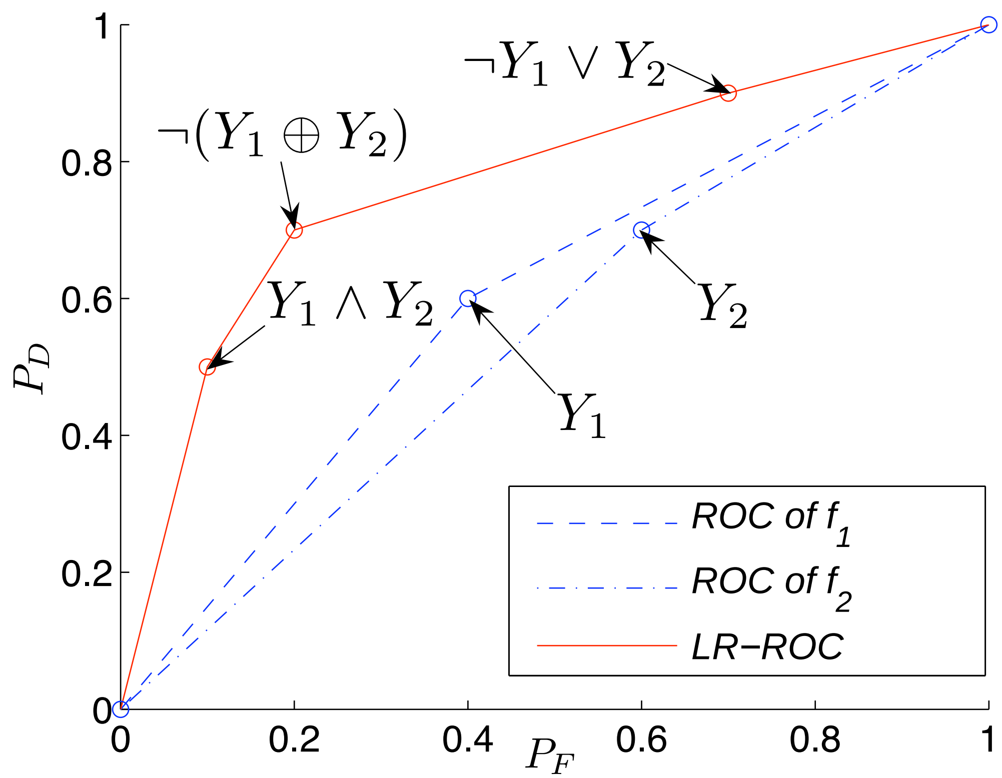
Y_2 \ Y_1	0	1
0	0.2	0.1
1	0.2	0.5

Class 0 (H_0)

Y_2 \ Y_1	0	1
0	0.1	0.3
1	0.5	0.1

$\ell(01)=2/5 < \ell(00)=2$

Example of the Likelihood-Ratio Test



Class 1 (H_1)

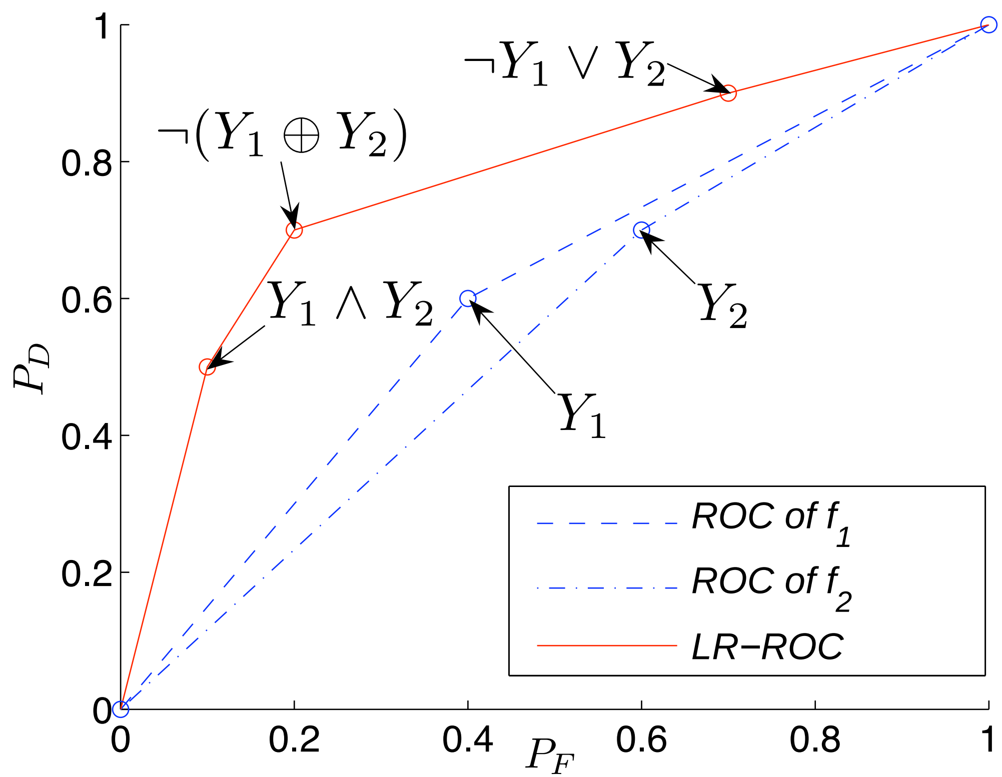
Y_2 \ Y_1	0	1
0	0.2	0.1
1	0.2	0.5

Class 0 (H_0)

Y_2 \ Y_1	0	1
0	0.1	0.3
1	0.5	0.1

$\ell(10)=1/3 < \ell(01)=2/5 < \ell(00)=2$

Example of the Likelihood-Ratio Test



Class 1 (H_1)

Y_2 \ Y_1	0	1
0	0.2	0.1
1	0.2	0.5

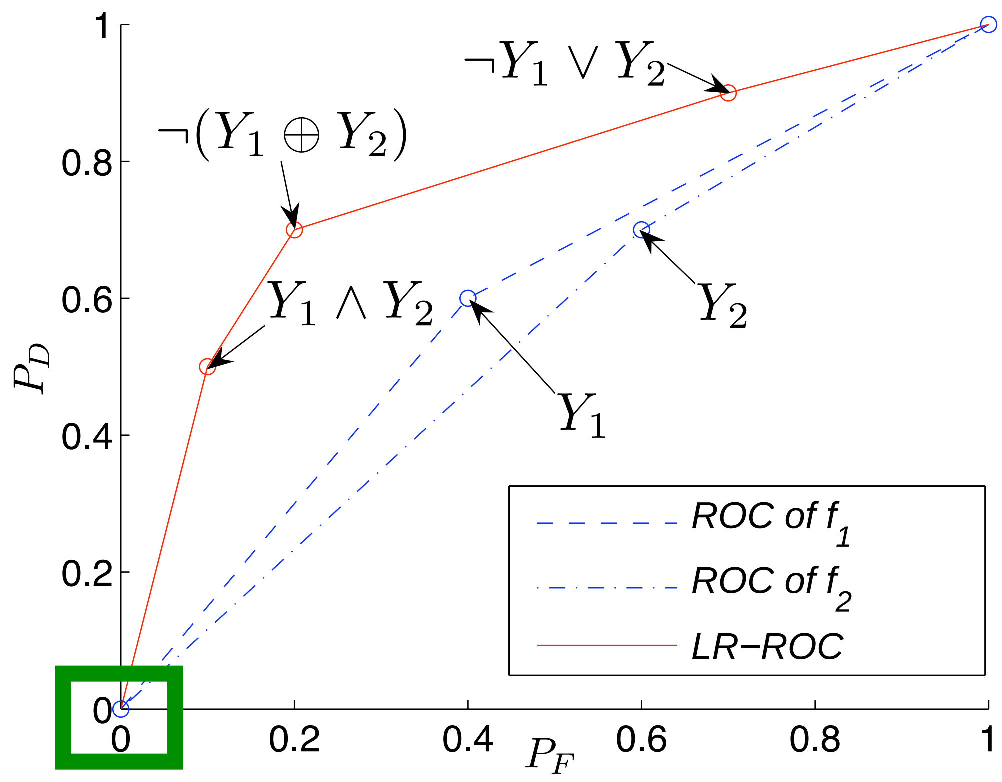
Class 0 (H_0)

Y_2 \ Y_1	0	1
0	0.1	0.3
1	0.5	0.1

$\ell(10)=1/3 < \ell(01)=2/5 <$

$\ell(00)=2 < \ell(11)=5$

Example of the Likelihood-Ratio Test



Class 1 (H_1)

Y_2 \ Y_1	0	1
0	0.2	0.1
1	0.2	0.5

Class 0 (H_0)

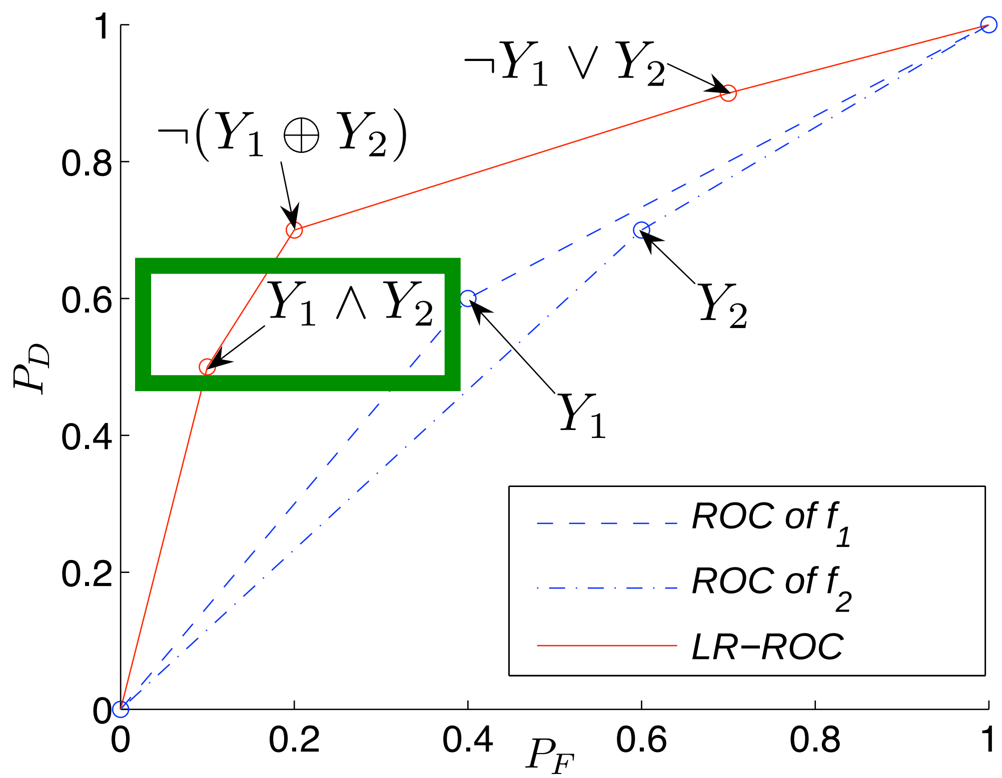
Y_2 \ Y_1	0	1
0	0.1	0.3
1	0.5	0.1

τ

$\ell(10) < \ell(01) < \ell(00) < \ell(11)$ |

$$Y_0 = 0$$

Example of the Likelihood-Ratio Test



Class 1 (H_1)

Y_2 \ Y_1	0	1
0	0.2	0.1
1	0.2	0.5

Class 0 (H_0)

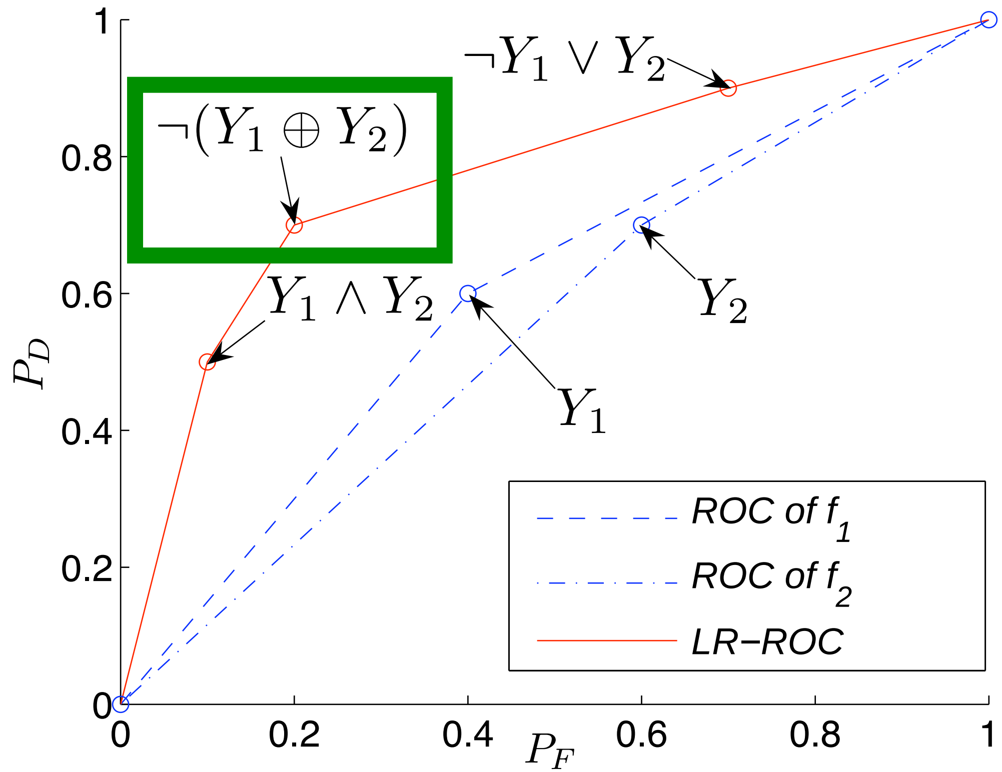
Y_2 \ Y_1	0	1
0	0.1	0.3
1	0.5	0.1

τ

$\ell(10) < \ell(01) < \ell(00) < \ell(11)$

$$Y_0 = Y_1 \wedge Y_2$$

Example of the Likelihood-Ratio Test



Class 1 (H_1)

Y_2 \ Y_1	0	1
0	0.2	0.1
1	0.2	0.5

Class 0 (H_0)

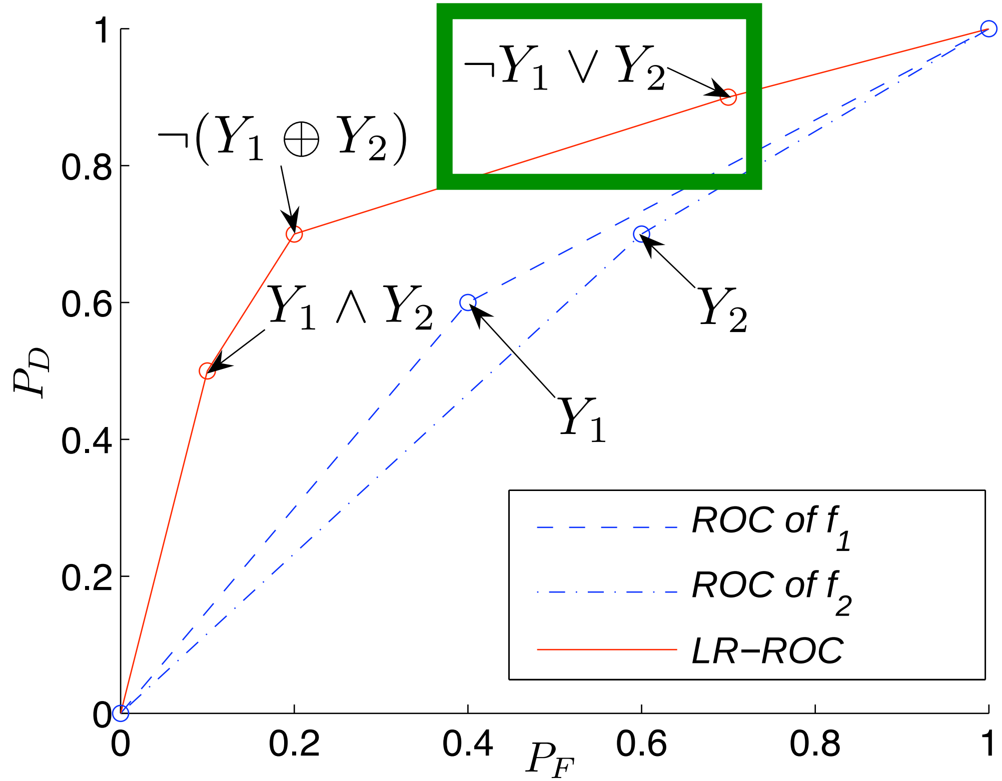
Y_2 \ Y_1	0	1
0	0.1	0.3
1	0.5	0.1

τ

$\ell(10) < \ell(01) < \ell(00) < \ell(11)$

$$
\begin{aligned}
Y_0 &= \bar{Y_1}\bar{Y_2} + Y_1 Y_2 \\
&= \neg(Y_1 \oplus Y_2)
\end{aligned}
$$

Example of the Likelihood-Ratio Test



Class 1 (H_1)

Y_2 \ Y_1	0	1
0	0.2	0.1
1	0.2	0.5

Class 0 (H_0)

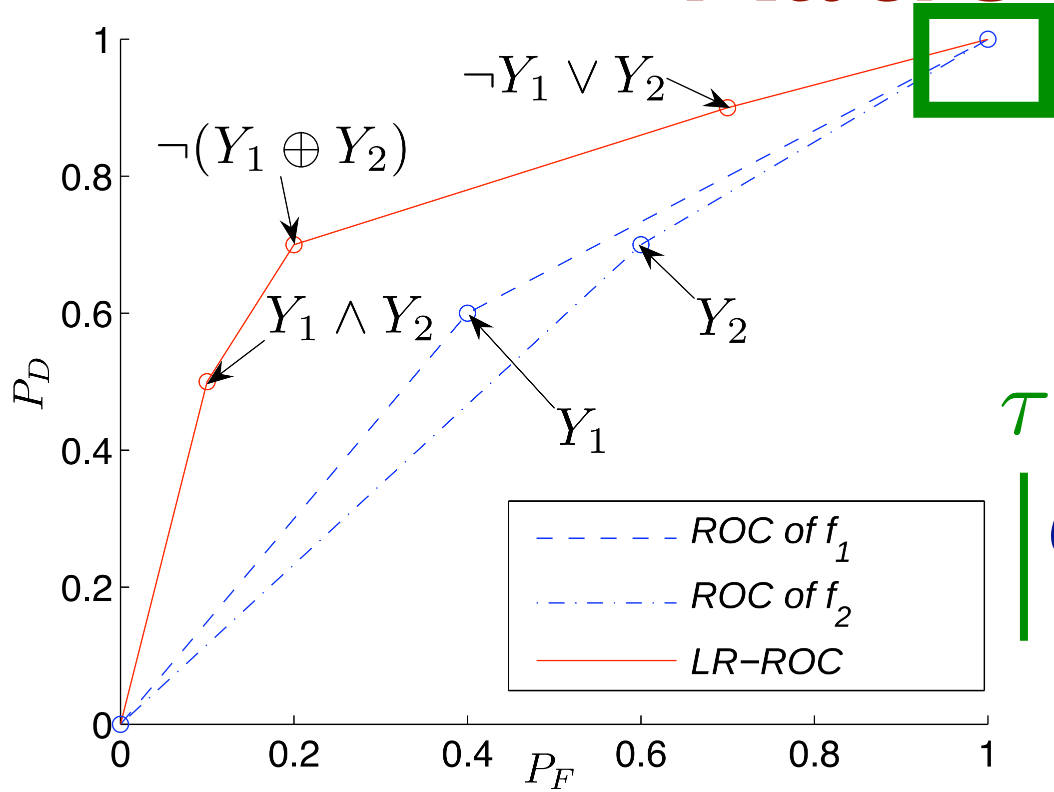
Y_2 \ Y_1	0	1
0	0.1	0.3
1	0.5	0.1

τ

$\ell(10) < \ell(01) < \ell(00) < \ell(11)$

$$
\begin{aligned}
Y_0 &= \bar{Y}_1 Y_2 + \bar{Y}_1 \bar{Y}_2 + Y_1 Y_2 \\
&= \neg Y_1 \vee Y_2
\end{aligned}
$$

Example of the Likelihood-Ratio Test



Class 1 (H_1)

Y_2 \ Y_1	0	1
0	0.2	0.1
1	0.2	0.5

Class 0 (H_0)

Y_2 \ Y_1	0	1
0	0.1	0.3
1	0.5	0.1

τ

$\ell(10)<\ell(01)<\ell(00)<\ell(11)$

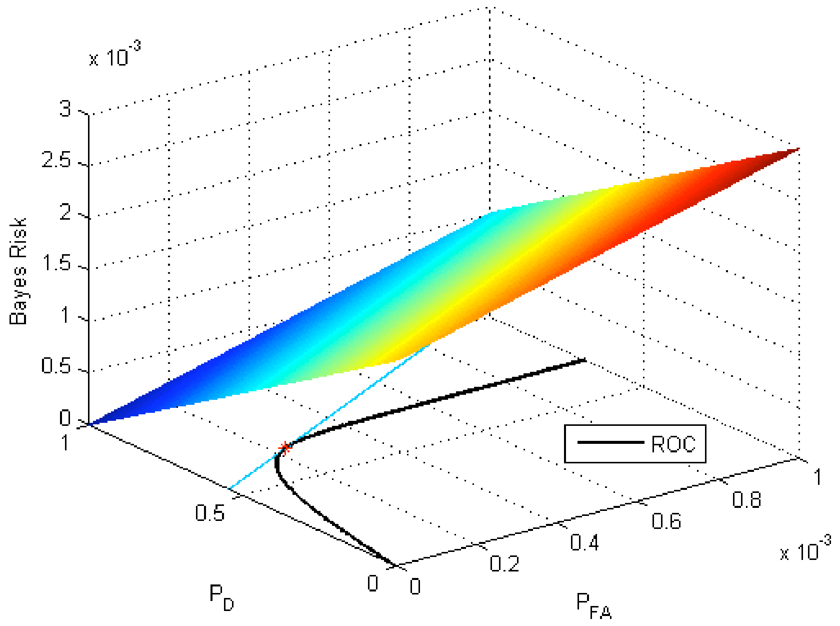
$$
\begin{aligned}
Y_0 &= Y_1\bar{Y_2} + \bar{Y_1}Y_2 + \bar{Y_1}\bar{Y_2} + Y_1Y_2 \\
&= 1
\end{aligned}
$$

Agenda

- Metric 1: Optimal ROC curve
- Metrics 2 & 3: Minimum cost and ranking
- Experiments
- Conclusions and Future Work

Metric 2: Expected Cost

- C_{01}=Cost of a false alarm
- C_{10}=Cost of a missed intrusion
- Expected Cost is a function of P_F and P_D
- The rule that minimizes the expected cost will lie in the ROC curve



Metric 3: Prioritization of Alerts

- The likelihood ratio is an estimate of the confidence for hypothesis H_1
- Example: $\ell(01) < \ell(10)$ =>
  - The alert given by $Y_1=1, Y_2=0$ should take priority over $Y_1=0, Y_2=1$.

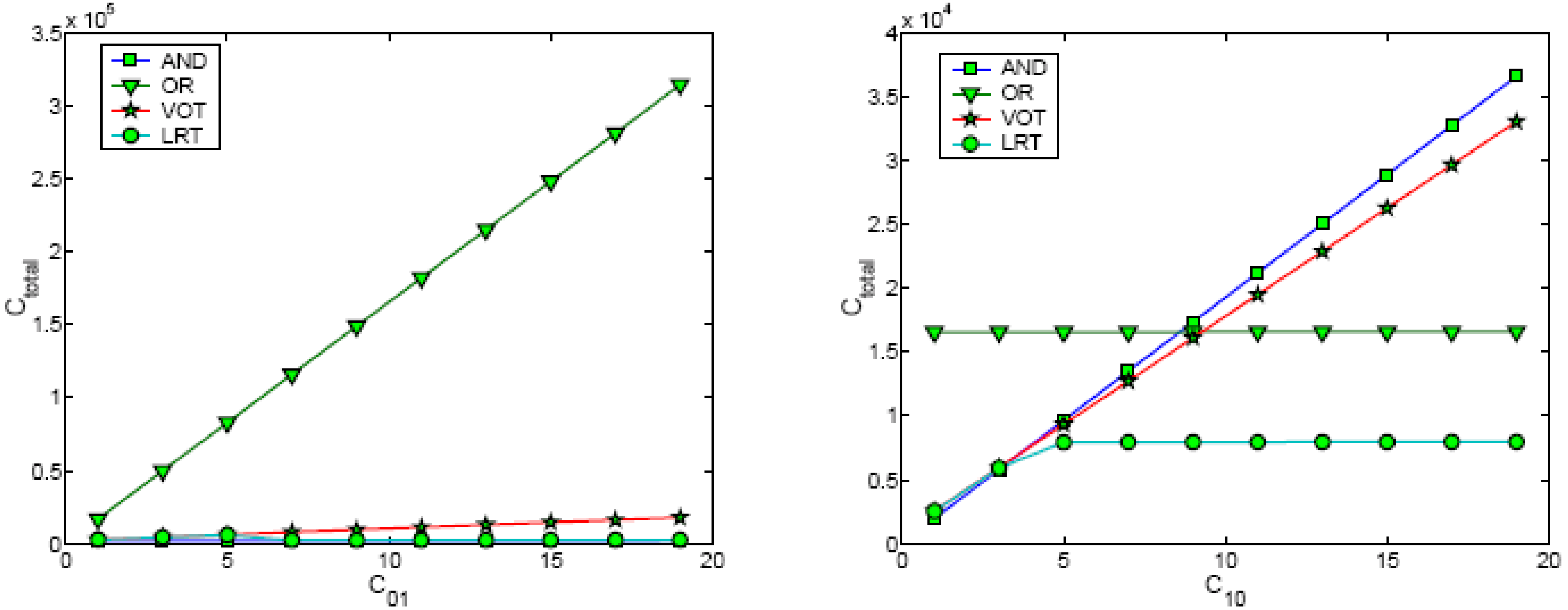
Agenda

- Metric 1: Optimal ROC curve
- Metrics 2 & 3: Minimum cost and ranking
- Experiments
- Conclusions and Future Work

Experiment Setup

- Dataset
  - Collected 30 minute HTTP trace (5 million packets) at College of Computing, Georgia Tech
  - Divided into two halves: training and testing set
  - Injected web attacks into testing set using tools, e.g., libwhisker (*base rate* 0.00082)
- Real-world IDSs
  - Snort (V2.3): signature based detection
  - PAYL: anomaly detector based on byte frequency within the payload
  - NetAD: modeling 48 attributes (48 bytes at fixed locations), summing up anomaly score based on byte frequency (within history, at the same location)

Experiment: Result



(a) Fix the cost of FN ($C_{10} = 1$) in all the cases, change the cost of FP (C_{01}).

(b) Fix the cost of FP ($C_{01} = 1$) in all the cases, change the cost of FN (C_{10}).

Experiment: Prioritization of Alerts

- Example: When PAYL raises an alarm alone, it should take precedence over when Snort and NetAD raise an alarm, but PAYL does not:

$$l(000) < l(001) < l(100) < l(101) < l(010) < l(011) < l(110) < l(111)$$

	Snort	PAYL	NetAD
P_D	0.016	0.99896	0.1037
P_F	0.0000237	0.00336	0.004

Snort = Y_1
PAYL=Y_2
NetAD=Y_3

Conclusions and Future Work

- We presented a theoretically sound and intuitive method for fusing alerts
- We generalized and improved previous work
- We plan to extend work to probabilistic IDS, and anomaly detectors