Towards an Information-Theoretic Framework for Analyzing Intrusion Detection Systems

Guofei Gu[1], Prahlad Fogla[1], David Dagon[1], Wenke Lee[1]
Boris Skoric[2]

[1]Georgia Institute of Technology

[2]Philips Research Laboratories, Netherlands

ESORICS'06

Outline

Outline

Problem Domain

- IDS is an essential component of the defense-in-depth security architecture.
- A number of IDSs have been proposed and developed in both industry and academic, which primarily focus on detection algorithms.
- Yet we still need to strengthen mathematical foundations and theoretic guidelines for IDS research
- Goal: to establish a mathematical foundation that can help us analyze and quantify the effectiveness of an IDS in both theory and practice

Outline

Our Contribution

- A uniform formal model of an IDS
- A fine-grained information-theoretic analysis on the IDS model
- A series of information-theoretic metrics that can quantitatively measure the effectiveness of an IDS and its components
- Our framework provides practical guidelines for fine-tuning, evaluation and design of IDSs.

Outline

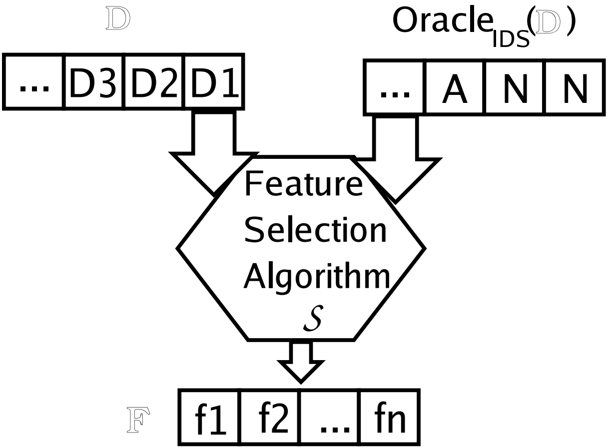
1. Motivation
   - The Basic Problem That We Studied
   - Our Contribution
2. An Information-Theoretic Framework for Analyzing IDSs
   - **Modeling an IDS**
   - Connection to Information Theory
   - Simplified Model Analysis
   - Implication
3. Experiments
   - Experiment Environment
   - Fine-tuning PAYL in static and dynamic situations
   - Improve IDS Design
4. Summary

Modeling an IDS

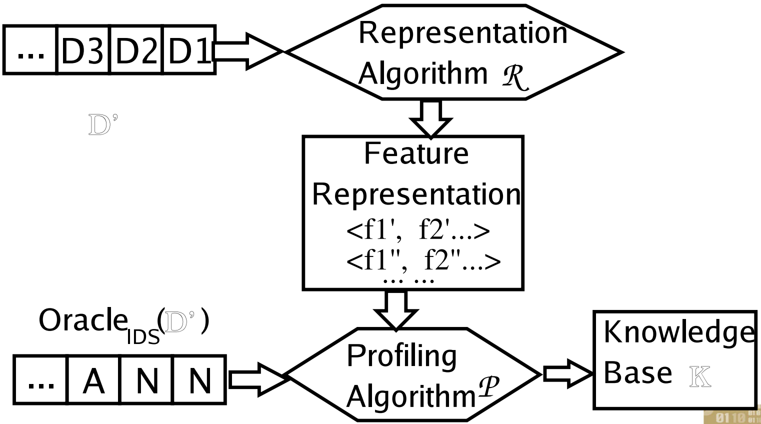
An IDS is an eight-tuple $(\mathbb{D}, \Sigma, \mathbb{F}, \mathbb{K}, \mathcal{S}, \mathcal{R}, \mathcal{P}, \mathcal{C})$

- $\mathbb{D}$: data source to examine and analyze, $\mathbb{D} = (D_1, D_2, ...)$
- Σ: data states, e.g., $\Sigma = \{N, A\}$
- $\mathbb{F}$: feature vector, $\mathbb{F} =< f_1, f_2, ... >$
- $\mathbb{K}$: knowledge base (profiles)
- $\mathcal{S}$: feature selection algorithm
- $\mathcal{R}$: data reduction and representation algorithm, $\mathcal{R} : \mathbb{D} \rightarrow \mathbb{F}$
- $\mathcal{P}$: profiling algorithm
- $\mathcal{C}$: classification algorithm, $\mathcal{C} : \mathbb{F} \rightarrow \Sigma$

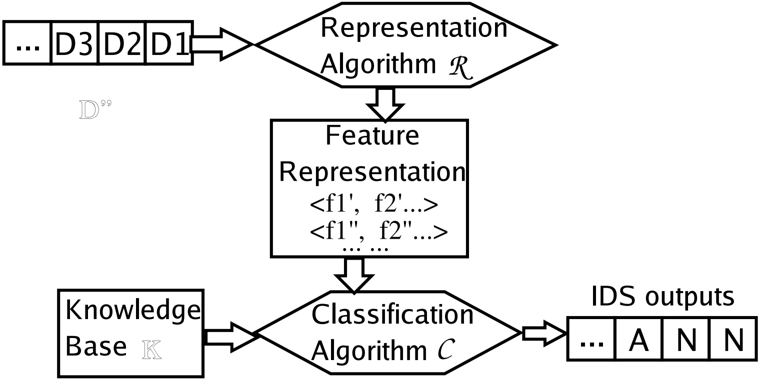
Feature selection procedure



Profiling/training procedure



Detection procedure



Outline

GEORGIA TECH INFORMATION SECURITY CENTER

Intrusion Detection Procedure Vs. Data Communication

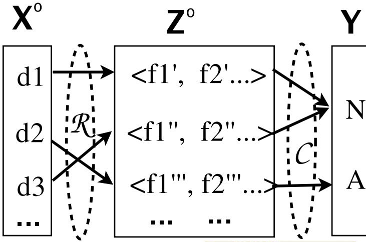
Similarity:

- Information processing and transmission
- $\mathcal{R}$ as an encoding algorithm
- $\mathcal{C}$ as a decoding algorithm

Difference:

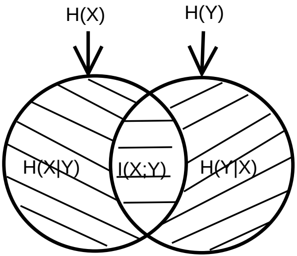
- We cannot enumerate all possible data input (source codes) and feature representation (code words) for an IDS
- IDS cannot keep a huge encoding/decoding table
- $\mathcal{R}, \mathcal{C}$ cannot guarantee errorless (information loss)

Original model



Information Theory Background

- Entropy $H(X)$: uncertainty of X
- Conditional entropy $H(X|Y)$: amount of uncertainty of X after Y is seen
- Mutual information $I(X,Y)$: reduction of uncertainty in X after Y is known (amount of information shared between X and Y)

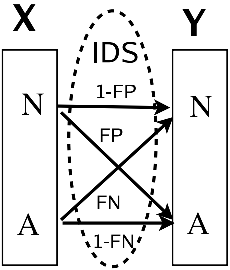


Outline

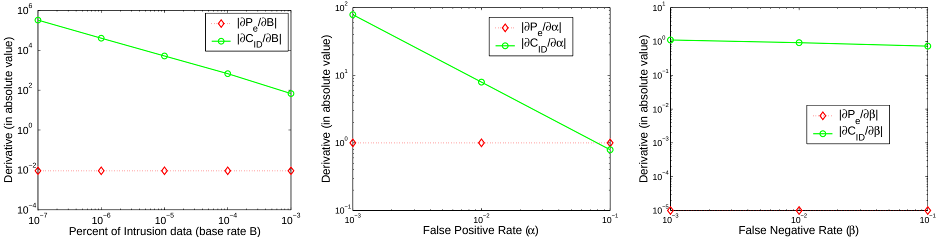
Abstract Model: Treat IDS as a black box

Intrusion detection capability $C_{ID} = \frac{I(X;Y)}{H(X)}$): How much (normalized) ground truth information an IDS can identify [Gu et al. ASIACCS'06]

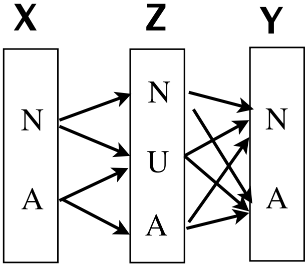
- take into account all aspects of detection capability
- an intrinsic measure of intrusion detection capability
- an objective trade-off between FP and FN (without involving subjective cost)
- yields a series of related information-theoretic metrics
- very sensitive and easy to demonstrate the effect of subtle changes of an IDS



Sensitivity analysis



Clustered Model



Cluster the feature representation vectors to only three states $\{N, U, A\}$

- Feature representation capability $C_R = \frac{I(X;Z)}{H(X)}$
- Classification information loss $L_C = \frac{I(X,Y;Z) - I(Y;Z)}{H(X)}$
- Theorem: $C_{ID} = C_R - L_C$
- Corollary: $C_{ID} \leq C_R$

Information Flow in an IDS

- Assume the original ground truth information is 1 (normalized)
- When the data reduction and representation algorithm $\mathcal{R}$ is applied, this information is reduced to C_R
- After the classification algorithm $\mathcal{C}$ is performed, there is further L_C amount of information loss
- The end result is C_{ID}, the overall capability of the IDS.

Outline

Implication

- With the help of C_{ID}, one can select the optimal operating point (where C_{ID} is maximized) for an IDS
- With the whole set of metrics, we provide a fine-grained analysis and quantification on the effectiveness of an IDS and its components
- We can identify whether and how the feature representation or classification algorithm is (the bottleneck) to be improved

Implication (cont.)

Importance of C_R

- Improve $\mathcal{R}$: full assembling, protocol parsing, normalization
- Improve feature set: to increase C_R (and carefully not increase L_C), e.g., context-aware signature

Outline

Experiment Environment

Experiment I: fine-tuning an IDS

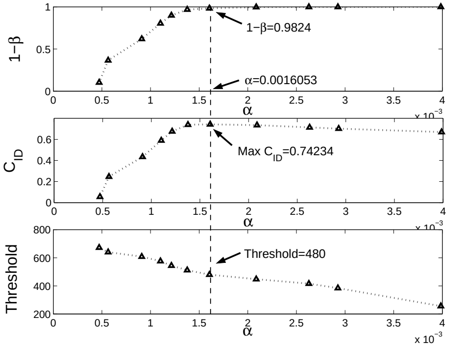
- PAYL as our IDS example
- CoC http traffic (7.5 million http request packets)
- Static and dynamic cases

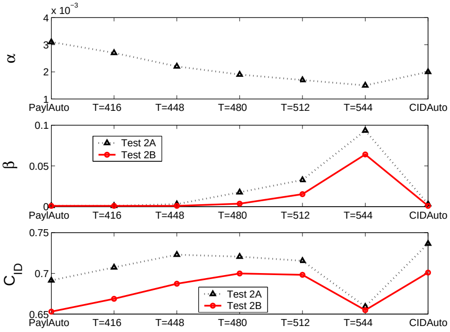
Experiment II: fine-grained evaluation and design improvement of IDSs

- Machine learning based IDSs
- 1998 DARPA Intrusion Detection Evaluation dataset (KDD cup 1999)

Outline

Fine-tuning PAYL in static and dynamic situations





Outline

Improve IDS Design

IDS1(with feature set 2 and C4.5), IDS2(with feature set 3 and Naive Bayes)

IDS	α	β	C_{ID}	C_R	L_C
IDS1	0.023699	0.079437	0.4002	0.8092	0.4090
IDS2	0.022577	0.10329	0.3875	0.9644	0.5769

IDS1*(after improving feature set), IDS2*(after improving classification algorithm)

IDS	α	β	C_{ID}	C_R	L_C
IDS1*	0.017609	0.089676	0.4258	0.9644	0.5386
IDS2*	0.017576	0.090374	0.4255	0.9644	0.5389

Improve IDS Design

IDS1(with feature set 2 and C4.5), IDS2(with feature set 3 and Naive Bayes)

IDS	α	β	C_{ID}	C_R	L_C
IDS1	0.023699	0.079437	0.4002	0.8092	0.4090
IDS2	0.022577	0.10329	0.3875	0.9644	0.5769

IDS1*(after improving feature set), IDS2*(after improving classification algorithm)

IDS	α	β	C_{ID}	C_R	L_C
IDS1*	0.017609	0.089676	0.4258	0.9644	0.5386
IDS2*	0.017576	0.090374	0.4255	0.9644	0.5389

GEORGIA TECH INFORMATION SECURITY CENTER

Summary

- A uniform formal model of an IDS
- A fine-grained information-theoretic analysis on the IDS model
- A series of information-theoretic metrics that can quantitatively measure the effectiveness of an IDS and its components
- Our framework provides practical guidelines for fine-tuning, evaluation and design of IDSs.

- Future work
  - IDS internal and external architecture study and design improvement.
  - More robust ways of applying the framework.

Thanks.

Any question or comment?