# Intrusion Detection Evaluation

Prof. Ravi Sandhu
Executive Director and Endowed Chair

Lecture 15

ravi.utsa@gmail.com
www.profsandhu.com

Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.

# IDS Categorization

| Property | IDS Type |
|---|---|
| Monitored platform | Host based |
| | Network based |
| | Hybrid |
| Attack detection method | Misuse based |
| | Anomaly based |
| | Hybrid |
| Deployment architecture | Nondistributed |
| | Distributed |

Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.
Table 1

# IDS Categorization

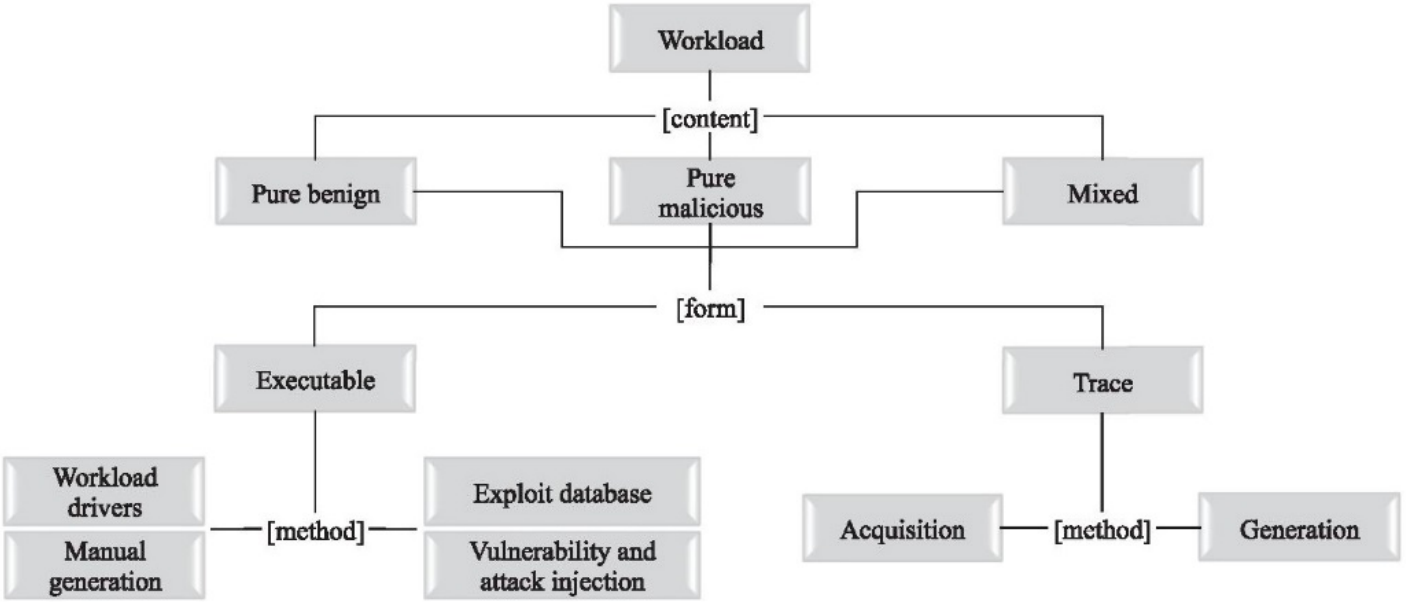
| Property | IDS Type |
|---|---|
| Monitored platform | Host based |
| | Network based |
| | Hybrid |
| Attack detection method | Misuse based |
| | Anomaly based |
| | Hybrid |
| Deployment architecture | Nondistributed |
| | Distributed |

Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.
Table 1

# Design Space

- Workloads
- Metrics
- Measurement methodology

# Workloads




Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.
Figure 1

# Workloads

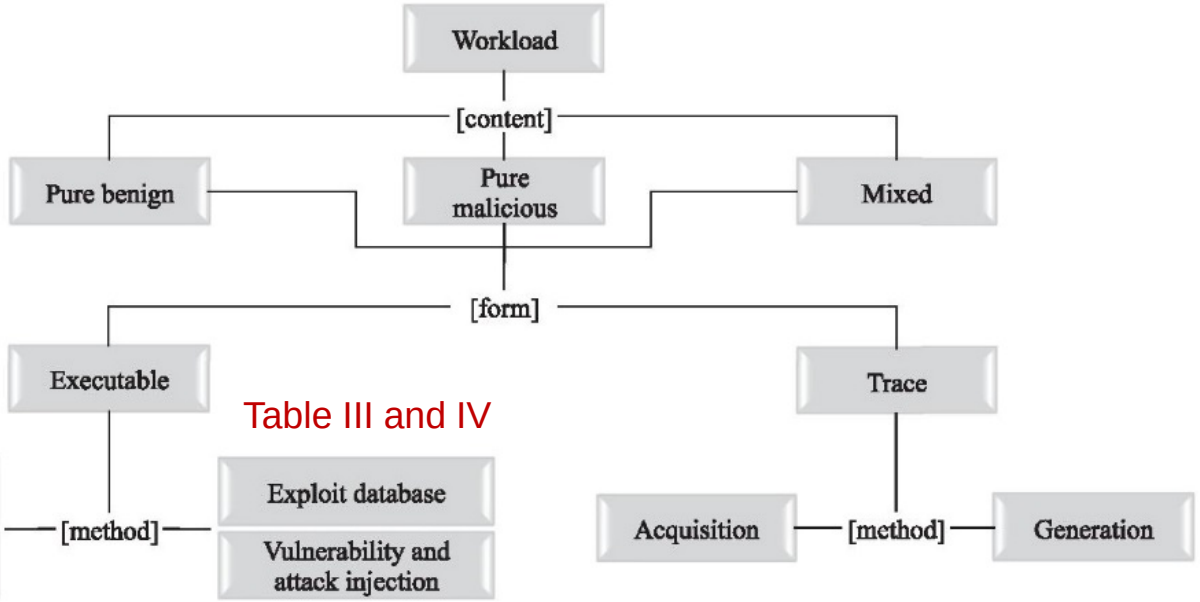



Table III and IV

Table II

Figure 2

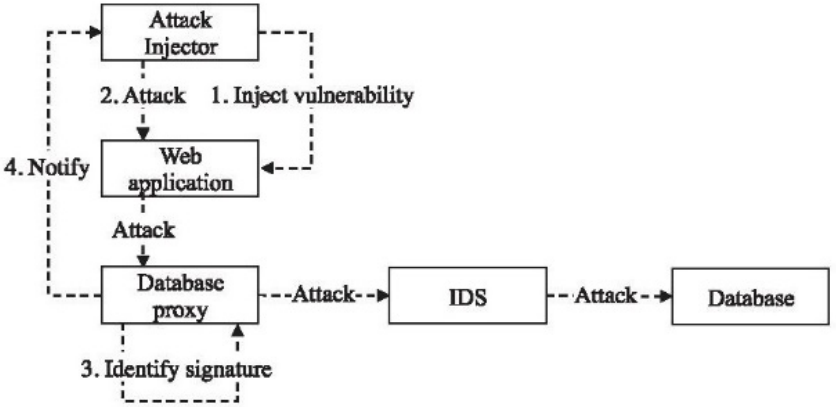
Publicly available traces
DARPA 98, 99, 00
KDD 99 (derivative)

Symantec onsite testing

Table V

Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.
Figure 1

# Vulnerability and Attack Injection




Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.
Figure 2

# Honeypots



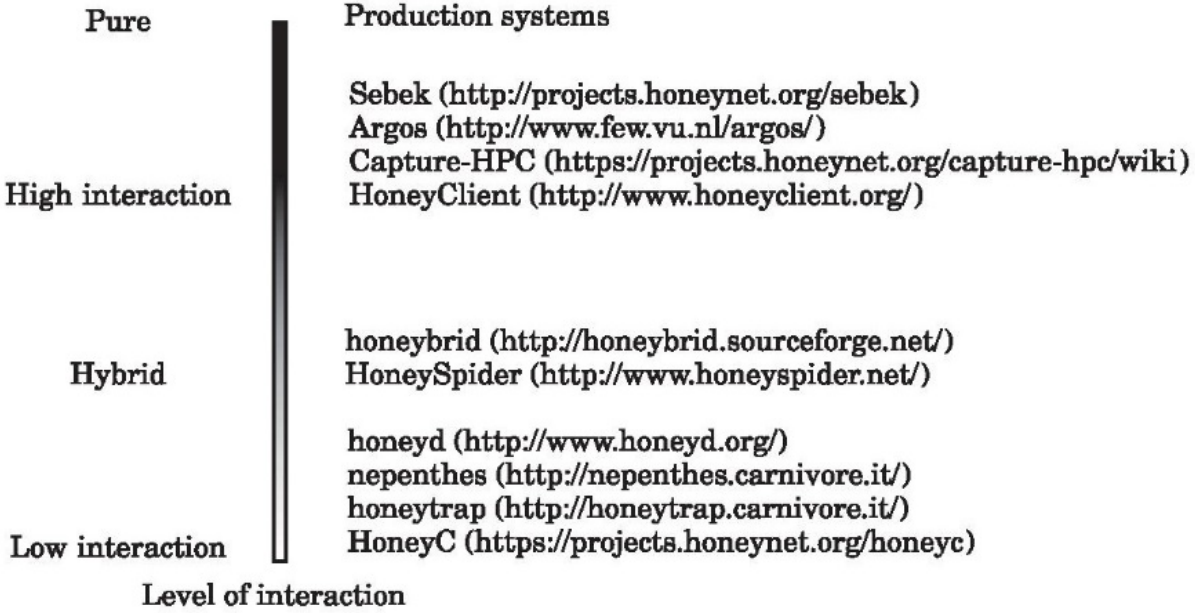
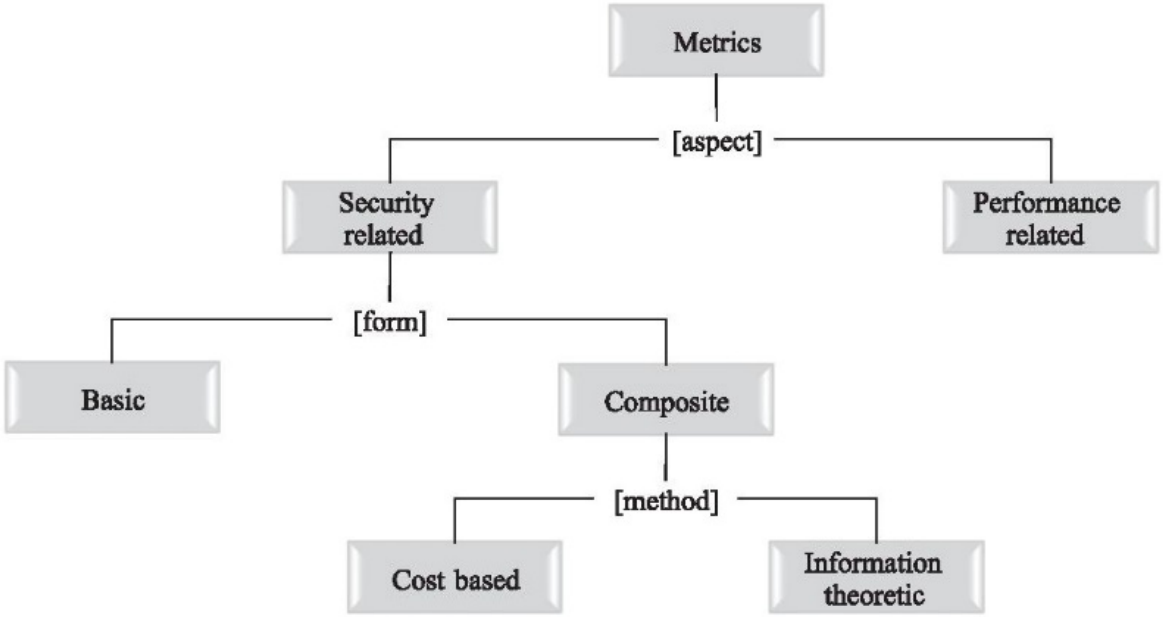



Fig. 3. Honeypots of different levels of interaction.

Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.
Figure 3

# Metrics






Not discussed in lecture

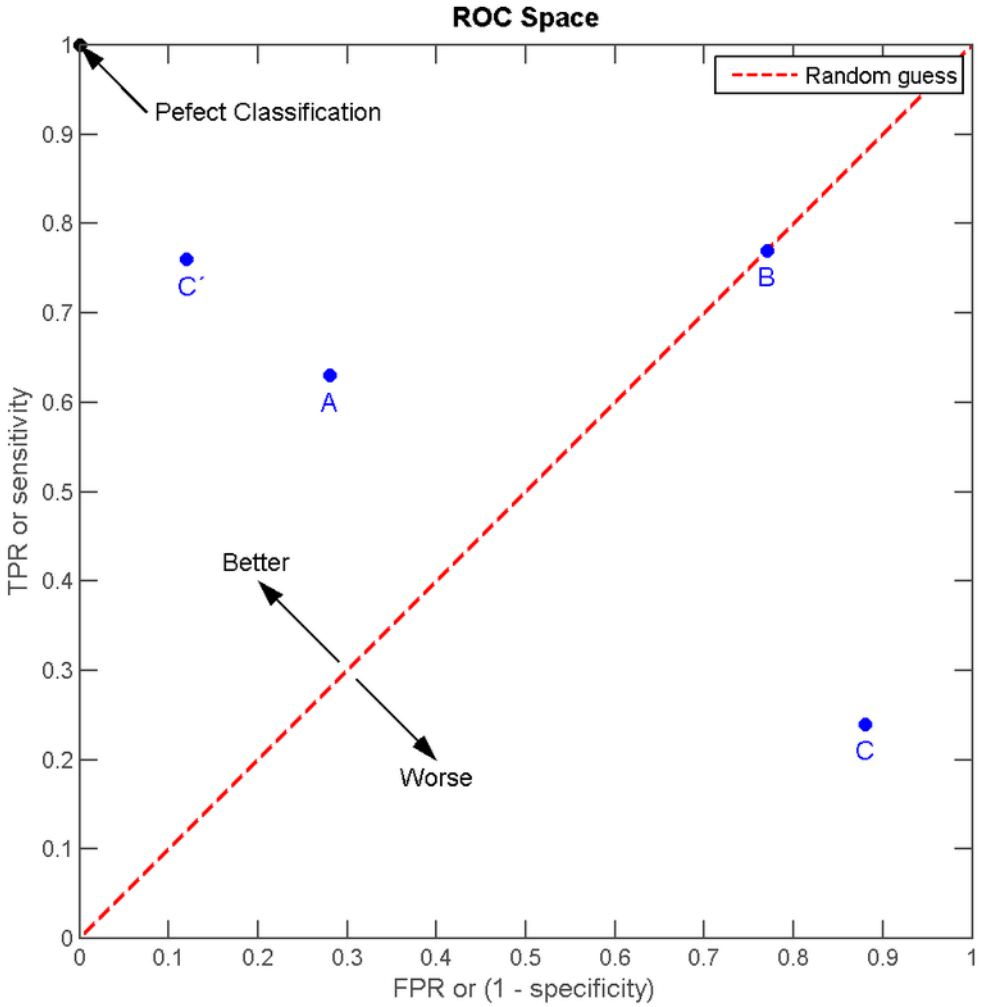
Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.
Figure 4

# Metrics: Basic

| | | |
|---|---|---|
| Basic | False-negative rate | $\beta = P(\neg A \mid I)$ |
| | True-positive rate | $1 - \beta = 1 - P(\neg A \mid I) = P(A \mid I)$ |
| | False-positive rate | $\alpha = P(A \mid \neg I)$ |
| | True-negative rate | $1 - \alpha = 1 - P(A \mid \neg I) = P(\neg A \mid \neg I)$ |
| Dependent on base rate | Positive predictive value | $P(I \mid A) = \frac{P(I)P(A \mid I)}{P(I)P(A \mid I) + P(\neg I)P(A \mid \neg I)}$ |
| | Negative predictive value | $P(\neg I \mid \neg A) = \frac{P(\neg I)P(\neg A \mid \neg I)}{P(\neg I)P(\neg A \mid \neg I) + P(I)P(\neg A \mid I)}$ |

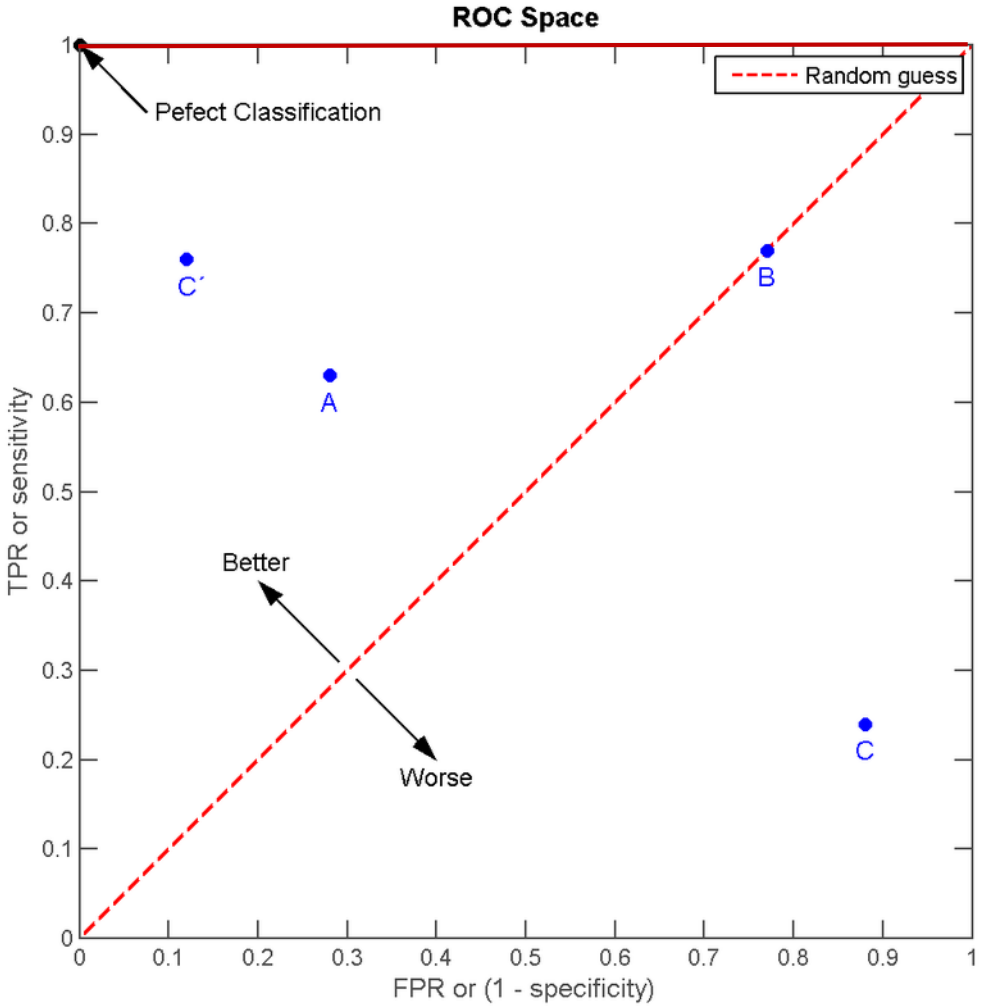
Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.
Figure 4

# Receiver Operating Curve (ROC)




https://en.wikipedia.org/wiki/Receiver_operating_characteristic

# Zero Reference Curve (ZRC)




https://en.wikipedia.org/wiki/Receiver_operating_characteristic

# ROC Limitations



- Intrusion detection is not a binary yes/no problem
- Unit of measurement is ambiguous
  - Flow versus packet
- Does not account for base rate P(I)

# Receiver Operating Curve (ROC)



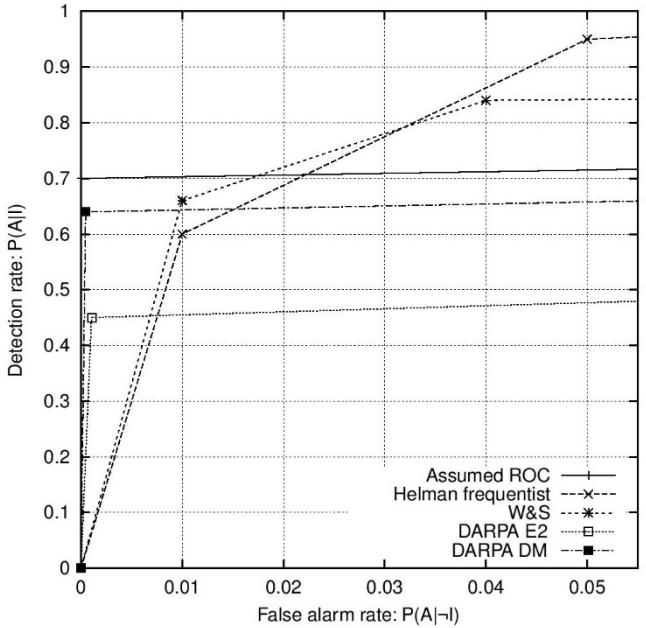



Fig. 2. ROC-curves for the "low performers".

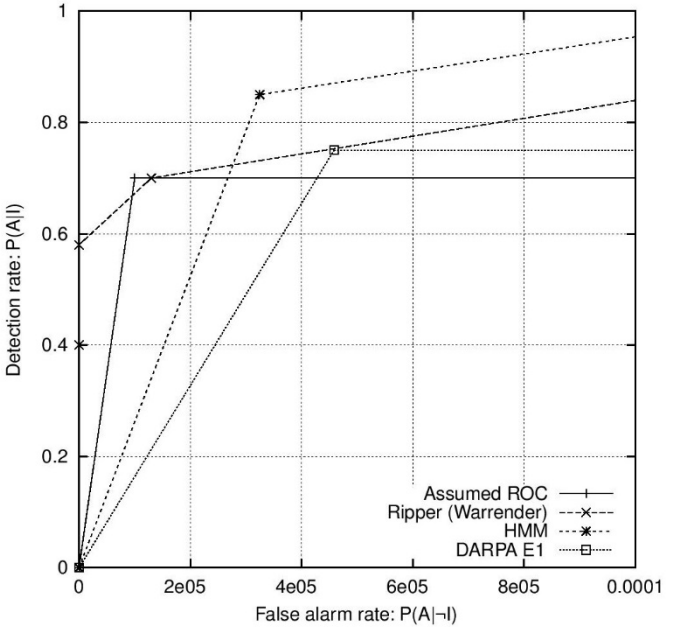



Fig. 3. ROC-curve for the "high performers".

Assumed ROC: fixes 1 point, 0.7 Detection rate, 0.00001 False alarm rate
Others are reported results from literature
All anomaly detectors are in Fig 2

Axelsson, Stefan. The base-rate fallacy and the difficulty of intrusion detection. ACM Transactions on Information and System Security (TISSEC) 3, no. 3 (2000): 186-205.
Figures 2 and 3

# Intrusion Detection Effectiveness



Table VII. Values of $1-\beta$, $PPV_{ID}$, $C_{exp}$, $C_{rec}$, and $C_{ID}$ for $IDS_1$ and $IDS_2$

| | | IDS$_1$ | | | | IDS$_2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $PPV_{ZRC}$ | $1-\beta$ | $PPV_{ID}$ | $C_{exp/rec}$ | $C_{ID}$ | $1-\beta$ | $PPV_{ID}$ | $C_{exp/rec}$ | $C_{ID}$ |
| 0.005 | 0,9569 | 0.9885 | 0,9565 | **0.016** | **0.9159** | 0.973 | 0,9558 | 0.032 | **0.8867** |
| 0.010 | 0,9174 | 0.99 | 0,9167 | 0.019 | 0.8807 | 0.99047 | 0,9167 | 0.019 | 0.8817 |
| 0.015 | 0,8811 | 0.9909 | 0,8801 | 0.022 | 0.8509 | 0.99664 | 0,8807 | **0.017** | 0.8635 |

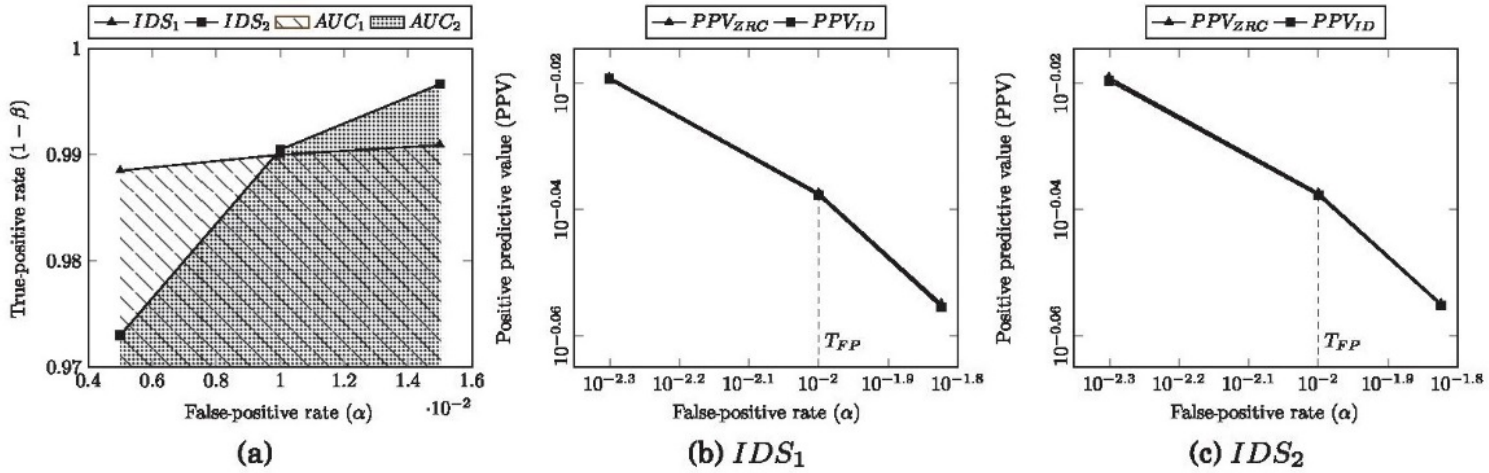



Fig. 5. IDS comparison with ROC curves (a) and the intrusion detection effectiveness metric (b, c).

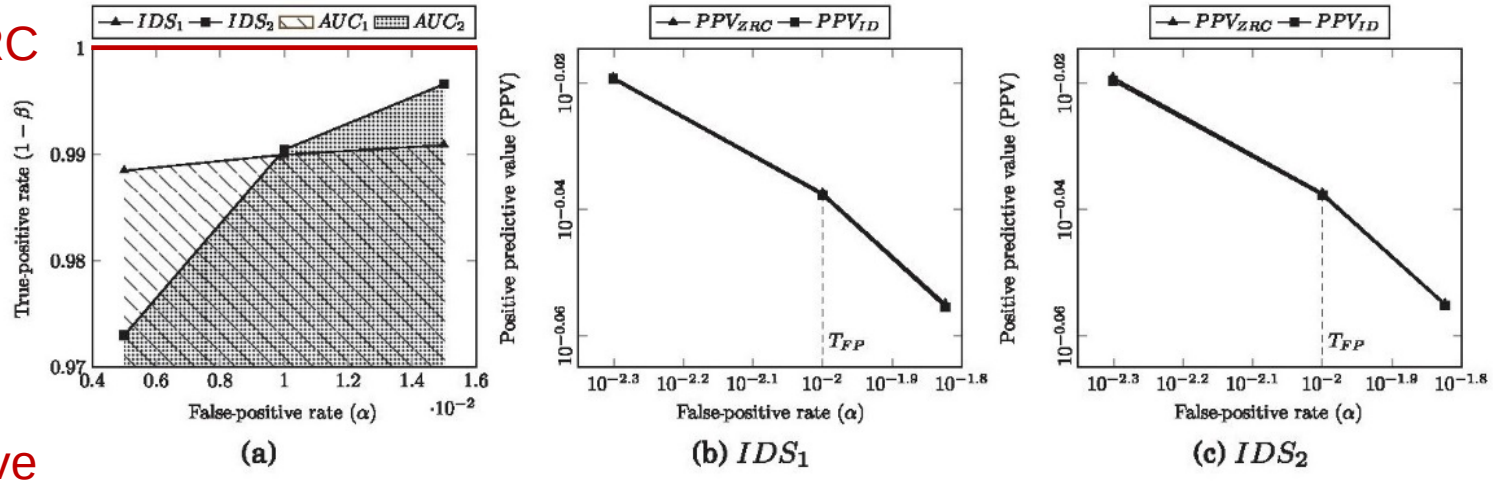
Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.

# Intrusion Detection Effectiveness

Table VII. Values of $1-\beta$, $PPV_{ID}$, $C_{exp}$, $C_{rec}$, and $C_{ID}$ for $IDS_1$ and $IDS_2$

| | | IDS$_1$ | | | | IDS$_2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $PPV_{ZRC}$ | $1-\beta$ | $PPV_{ID}$ | $C_{exp/rec}$ | $C_{ID}$ | $1-\beta$ | $PPV_{ID}$ | $C_{exp/rec}$ | $C_{ID}$ |
| 0.005 | 0,9569 | 0.9885 | 0,9565 | **0.016** | **0.9159** | 0.973 | 0,9558 | 0.032 | **0.8867** |
| 0.010 | 0,9174 | 0.99 | 0,9167 | 0.019 | 0.8807 | 0.99047 | 0,9167 | 0.019 | 0.8817 |
| 0.015 | 0,8811 | 0.9909 | 0,8801 | 0.022 | 0.8509 | 0.99664 | 0,8807 | **0.017** | 0.8635 |

ZRC




Compare area under ROC curve

Fig. 5. IDS comparison with ROC curves (a) and the intrusion detection effectiveness metric (b, c).

Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.

# Intrusion Detection Effectiveness

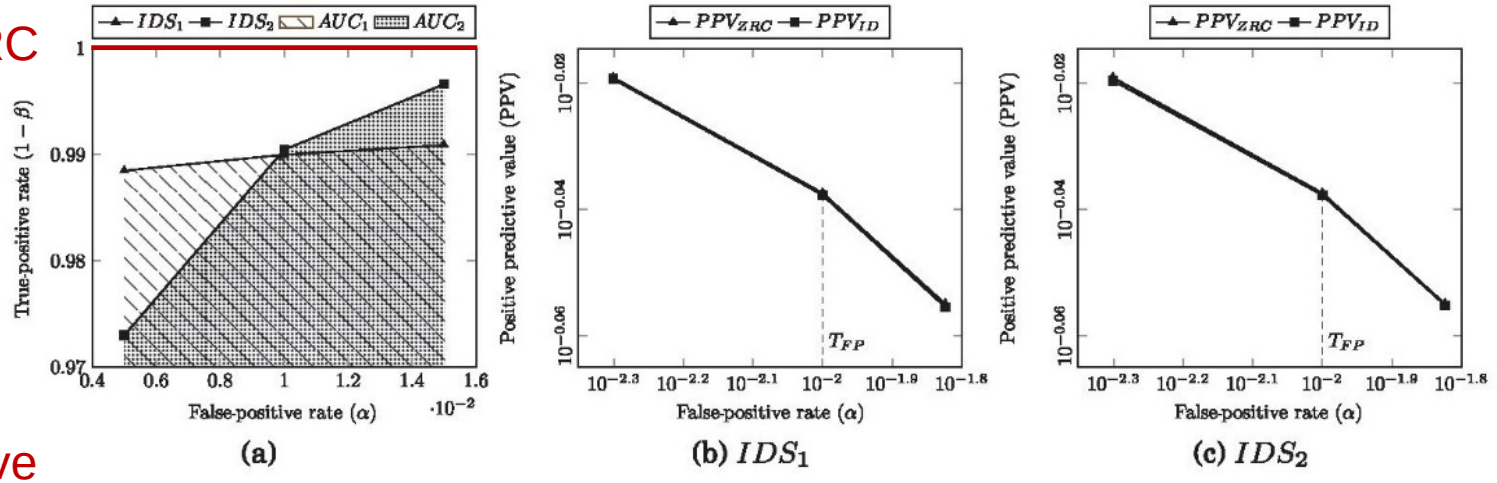


Assumes Base rate, P(I) = 0.1

Table VII. Values of $1-\beta$, $PPV_{ID}$, $C_{exp}$, $C_{rec}$, and $C_{ID}$ for $IDS_1$ and $IDS_2$

| | | IDS$_1$ | | | | IDS$_2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $PPV_{ZRC}$ | $1-\beta$ | $PPV_{ID}$ | $C_{exp/rec}$ | $C_{ID}$ | $1-\beta$ | $PPV_{ID}$ | $C_{exp/rec}$ | $C_{ID}$ |
| 0.005 | 0,9569 | 0.9885 | 0,9565 | **0.016** | **0.9159** | 0.973 | 0,9558 | 0.032 | **0.8867** |
| 0.010 | 0,9174 | 0.99 | 0,9167 | 0.019 | 0.8807 | 0.99047 | 0,9167 | 0.019 | 0.8817 |
| 0.015 | 0,8811 | 0.9909 | 0,8801 | 0.022 | 0.8509 | 0.99664 | 0,8807 | **0.017** | 0.8635 |

ZRC




Compare area under ROC curve

Fig. 5. IDS comparison with ROC curves (a) and the intrusion detection effectiveness metric (b, c).

$T_{FP}$: max acceptable false positive rate
Compare area difference between $PPV_{ZRC}$ and $PPV_{IDS}$ up to $T_{FP}$

Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.

# Metrics: Cost-Based

These $p_1$ $p_2$ $p_3$ are different, apply to false alert filter

$C_\alpha$: cost of false positive
$C_\beta$: cost of false negative
$C = C_\beta/C_\alpha$

$p_1 = P(A)$
$p_2 = P(I|A)$
$p_3 = P(I|\neg A)$

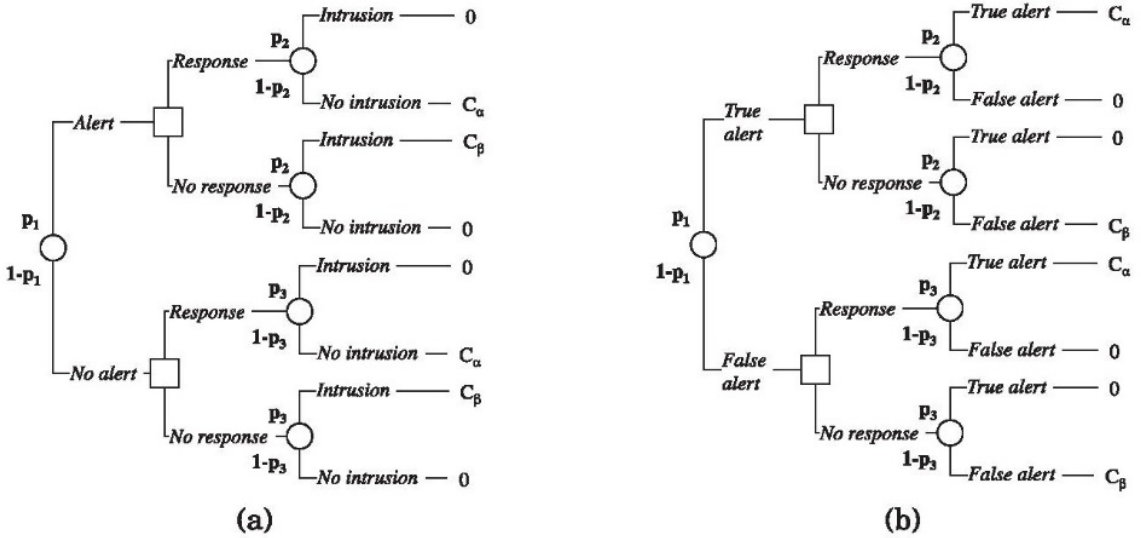



Fig. 6. Decision tree for calculating expected cost (a) and relative expected cost (b).

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

$C_{rec} = C\beta B + \alpha(1-B)$

Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.
Figure 6

# Metrics: Cost-Based



Table VII. Values of $1-\beta$, $PPV_{ID}$, $C_{exp}$, $C_{rec}$, and $C_{ID}$ for $IDS_1$ and $IDS_2$

| | | IDS$_1$ | | | | IDS$_2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $PPV_{ZRC}$ | $1-\beta$ | $PPV_{ID}$ | $C_{exp/rec}$ | $C_{ID}$ | $1-\beta$ | $PPV_{ID}$ | $C_{exp/rec}$ | $C_{ID}$ |
| 0.005 | 0,9569 | 0.9885 | 0,9565 | **0.016** | **0.9159** | 0.973 | 0,9558 | 0.032 | **0.8867** |
| 0.010 | 0,9174 | 0.99 | 0,9167 | 0.019 | 0.8807 | 0.99047 | 0,9167 | 0.019 | 0.8817 |
| 0.015 | 0,8811 | 0.9909 | 0,8801 | 0.022 | 0.8509 | 0.99664 | 0,8807 | **0.017** | 0.8635 |

Assumptions:
B = 0.1
C = 10
α, β same for base IDS and its false alarm filter

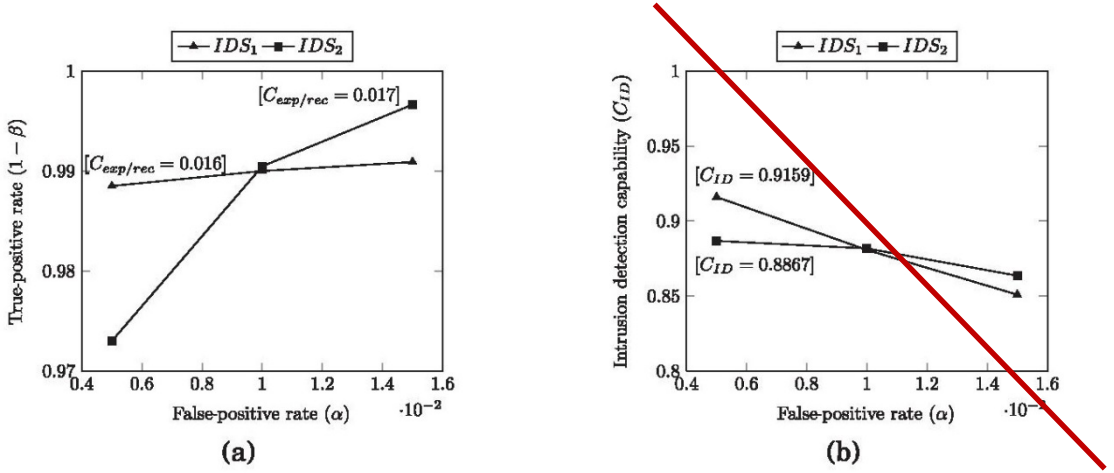



Fig. 7. IDS comparison with the expected cost and relative expected cost metric (a) and the intrusion detection capability metric (b).

Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.

# Measurement Methodology

Table VIII. IDS Evaluation Design Space: Measurement Methodology

| IDS Property | Workloads [Content] | Metrics [Aspect] | Metrics [Form] |
|---|---|---|---|
| **Attack Detection Related** | | | |
| Attack detection accuracy | Mixed | Security related | Basic, composite |
| Attack coverage | Pure malicious | Security related | Basic |
| Resistance to evasion techniques | Pure malicious, mixed | Security related | Basic |
| Attack detection and reporting speed | Mixed | Performance related | n/a |
| **Resource Consumption Related** | | | |
| CPU consumption<br>Memory consumption<br>Network consumption | Pure benign | Performance related | n/a |
| Performance overhead | Pure benign | Performance related | n/a |
| Workload processing capacity | Pure benign | Performance related | n/a |

**Definitions of IDS Properties**

| IDS Property | Definition |
|---|---|
| Attack detection accuracy | The attack detection accuracy of an IDS in the presence of mixed workloads. |
| Attack coverage | The attack detection accuracy of an IDS in the presence of attacks without any background benign activity. |
| Performance overhead | The overhead incurred by an IDS on the system and/or network environment where it is deployed. Under overhead, we understand performance degradation of users' tasks/operations caused by (a) consumption of system resources (e.g., CPU, memory) by the IDS and/or (b) interception and analysis of the workloads of users' tasks/operations (e.g., network packets) by the IDS. |
| Workload processing capacity | The rate of arrival of workloads to an IDS for processing in relation to the amount of workloads that the IDS discards (i.e., does not manage to process). For instance, in the context of network-based IDSes, capacity is normally measured as the rate of arrival of network packets to an IDS over time in relation to the amount of discarded packets over time. The capacity of an IDS may also be defined as the maximum workload processing rate of the IDS such that there are no discarded workloads. |

Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.

# Case Study: Snort

Table X. Attack Coverage of Snort

| Targeted Vulnerability (CVE ID) | Platform | Detected |
|---|---|---|
| CVE-2011-3192 | Apache | x |
| CVE-2010-1870 | Apache Struts | ✓ |
| CVE-2012-0391 | Apache Struts | x |
| CVE-2013-2251 | Apache Struts | x |
| CVE-2013-2115/CVE-2013-1966 | Apache Struts | ✓ |
| CVE-2009-0580 | Apache Tomcat | x |
| CVE-2009-3843 | Apache Tomcat | x |
| CVE-2010-2227 | Apache Tomcat | x |

✓, detected; x, not detected.

True positive rate = 2/8 = 0.25

Table XI. Resistance to Evasion Techniques of Snort

| Evasion Technique | Targeted Vulnerability (CVE ID) | |
|---|---|---|
| | CVE-2010-1870 | CVE-2013-2115/CVE-2013-1966 |
| HTTP::uri_use_backslashes | ✓ | ✓ |
| HTTP::uri_fake_end | ✓ | ✓ |
| HTTP::pad_get_params | ✓ | x |
| HTTP::uri_fake_params_start | ✓ | ✓ |
| HTTP::uri_encode_mode (u-random; hex-random) | ✓ | x |
| HTTP::pad_method_uri_count | ✓ | ✓ |
| HTTP::method_random_valid | ✓ | x |
| HTTP::header_folding | ✓ | ✓ |
| HTTP::uri_full_url | ✓ | ✓ |
| HTTP::pad_post_params | ✓ | x |
| HTTP::uri_dir_fake_relative | ✓ | ✓ |
| HTTP::pad_uri_version_type (apache; tab) | ✓ | ✓ |
| HTTP::uri_dir_self_reference | ✓ | ✓ |
| HTTP::method_random_case | ✓ | ✓ |

✓, detected; x, not detected.

True positive rate = 24/28 = 0.85

Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.

# Case Study: Snort



Table XII. Attack Detection Accuracy of Snort: Basic Metrics (seconds=120)

| Configuration | Metrics | | | |
|---|---|---|---|---|
| | $\alpha$ | $1-\beta$ | PPV | NPV |
| count=6 | 0.0008 | 0.333 | 0.9788 | 0.9310 |
| count=5 | 0.0011 | 0.416 | 0.9768 | 0.9390 |
| count=4 | 0.0013 | 0.5 | 0.9771 | 0.9473 |
| count=3 | 0.0017 | 0.624 | 0.9761 | 0.9598 |
| count=2 | 0.0024 | 0.833 | 0.9747 | 0.9817 |
| Default configuration | 0.0026 | 0.958 | 0.9762 | 0.9953 |

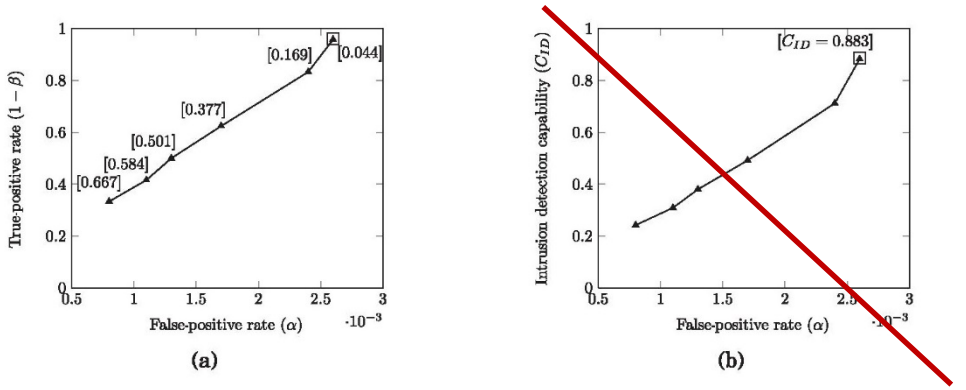



Fig. 8. Attack detection accuracy of Snort: composite metrics. ROC curve and estimated costs (a) and $C_{ID}$ curve (b) (□ marks an optimal operating point).

Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.

# Case Study: Snort



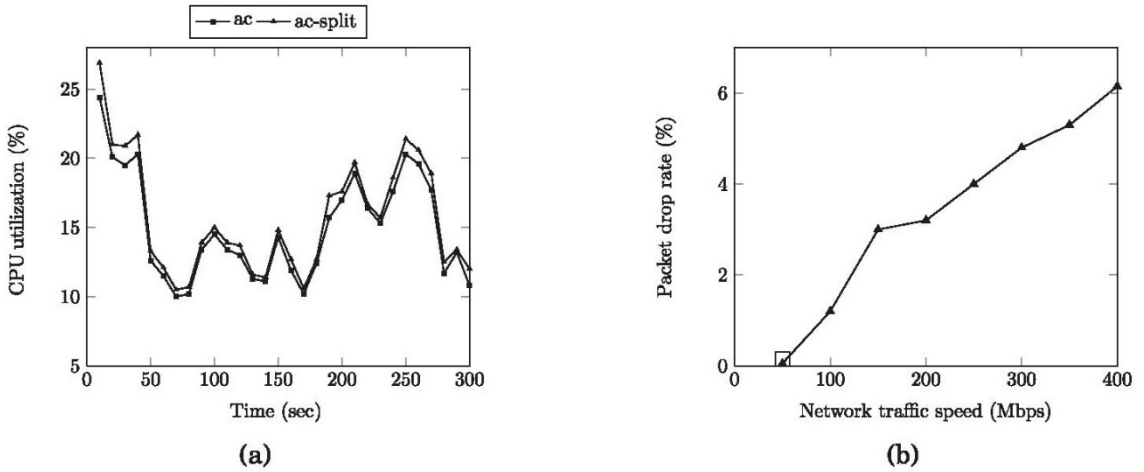



Fig. 9. CPU consumption of Snort (a) and packet drop rate of Snort (b) (□ marks the data point whose $x$ value is the network traffic speed that corresponds to the maximum workload processing rate of Snort such that there are no discarded workloads).

Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.

# Measurement Methodology



Table XIII. Summarizing Overview of Common Trends, Recommendations, and Key Best Practices

| IDS Property | |
|---|---|
| Attack detection accuracy<br>Attack coverage | These properties are evaluated for IDSes of all types. • The dated DARPA and KDD-99 Cup datasets represent at this time standard workloads for comparing novel anomaly-based IDSes with their past counterparts. • For the sake of representativeness, evaluate an IDS using not the DARPA or the KDD-99 Cup dataset but workloads that contain current attacks. • Attack detection rates of current IDSes vary greatly—that is, between 8% and 97%, measures that depend on the configurations of the tested IDSes and the applied evaluation methodologies. |
| Attack detection and reporting speed | This property is normally evaluated for distributed IDSes—it is best evaluated by measuring the time needed for the IDS to converge to a state in which all of its nodes, or the designated nodes, are notified of an ongoing attack. • Attack detection delays up to 3 seconds are considered acceptable. |
| Resistance to evasion techniques | This property is often not evaluated, as it is considered of limited practical importance. • Consider evaluating this property since a single successful IDS evasion attack poses the danger of a high-impact intrusion. • Metasploit is deemed the optimal tool for executing IDS evasive attacks, which is required for evaluating this property. • Many current IDSes are vulnerable to temporally crafted attacks. |
| Resource consumption related | These properties are typically evaluated for IDSes deployed in resource-constrained environments. • Network consumption in particular is often evaluated for distributed IDSes. • The resource consumption of a distributed IDS operating in wireless ad hoc networks is typically evaluated to measure the power consumption of its nodes—this is best performed by using a model that estimates power consumption based on resource consumption measurements. |
| Performance overhead | This property is normally evaluated for host-based IDSes. • Performance overhead is evaluated by executing tasks twice, once with the tested IDS being inactive and once with it being active. • This property is normally evaluated using workloads in executable form generated by workload drivers—workload drivers enable the straightforward generation of live customized workloads in a repeatable manner. • Overheads under 10%, relative to the execution time of tasks measured when the tested IDS is inactive, are generally considered acceptable. |
| Workload processing capacity | This property is normally evaluated for network-based IDSes that monitor high-rate workloads. • This property is best evaluated using traces or workload drivers, as they allow for the generation of workloads at user-defined speeds. • Evaluate the capacity of an IDS together with its resource consumption—this enables one to observe how resource consumption scales as workload intensity increases. |

Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A. and Payne, B.D., 2015. Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), p.12.

# Future Directions



- High speed IDSs
- IDSs for virtualized environments (e.g., cloud)
- IDSs for detecting APTs (advanced persistent threats)
- IDSs for detecting zero day attacks