IDS Detection Performance

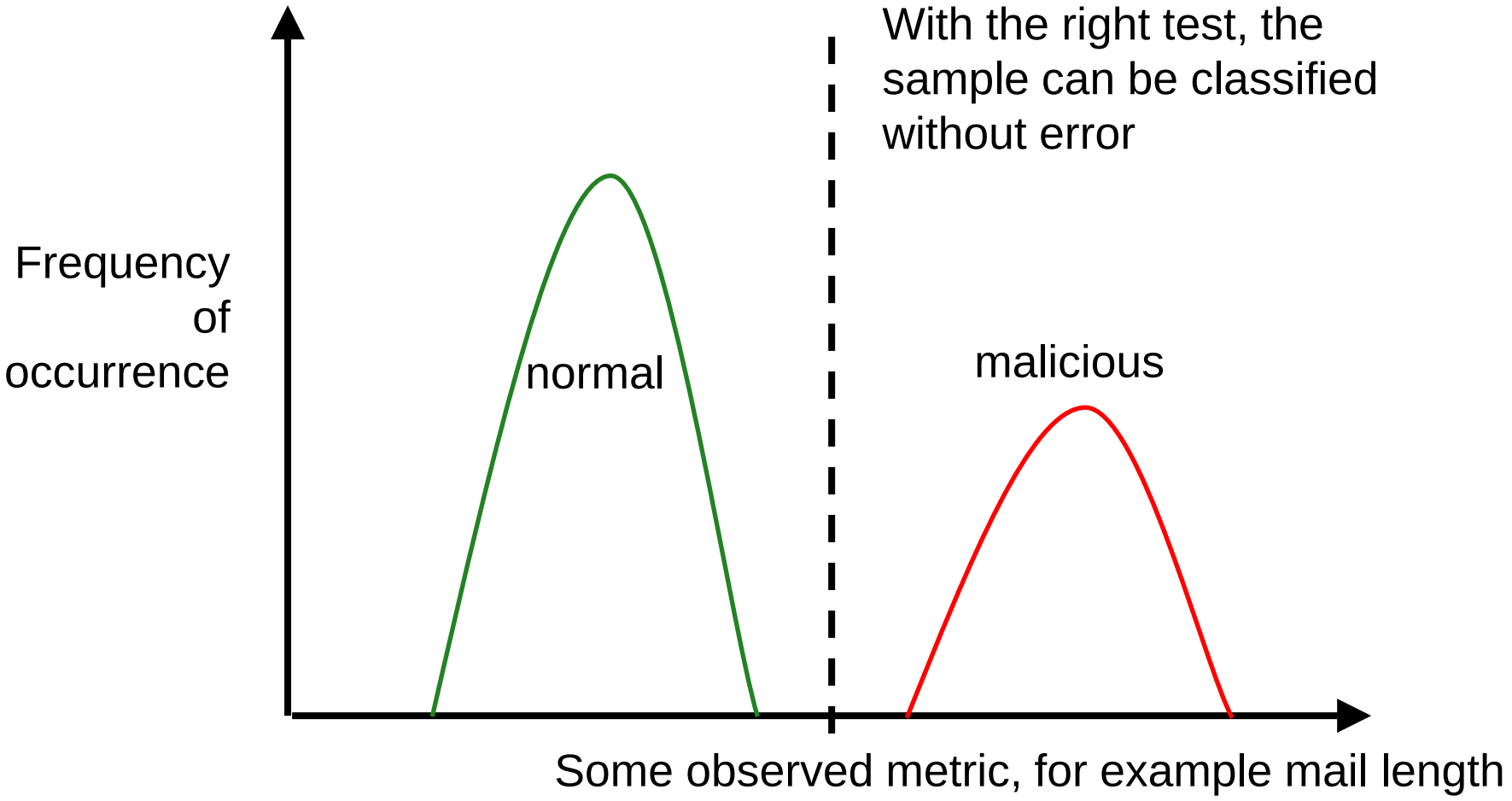
Ramin Sadre

Classification

- An IDS is a (binary) classifier: It takes some sample and has to decide whether it is normal or malicious
  - Input can be a log file entry, network packet, a function call,…
  - Input can be also a system state, a sequence of packets, a sequence of function calls,…

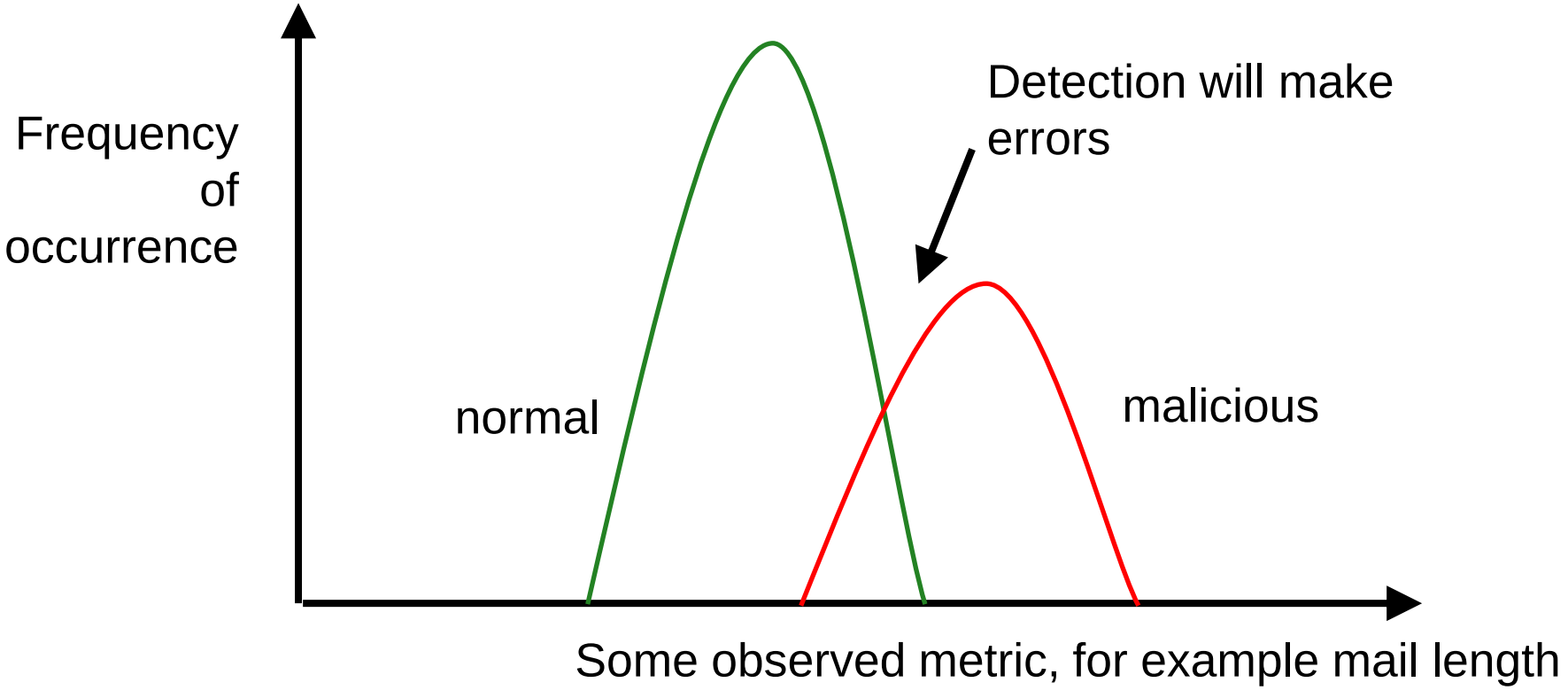
Classification (2)

- Ideally, normal samples are very different from malicious samples



Detection error

- In reality, normal and malicious samples can overlap for a chosen metric
- Of course, you can choose another metric, but 0% error is in general impossible



Ground truth

- How do you know whether your IDS makes mistakes?
- Important for
  - Developers of new IDS
  - Users of IDS

- You need a *ground truth:*

  An input dataset where each sample is *labeled* as normal or malicious
- Very hard to get good ground truth datasets
  - Must contain all possible attacks. What about 0-day attacks?
  - Must contain realistic normal data, otherwise the test might become too easy

Performance of an IDS

- To test an IDS you run it on the ground truth dataset and compare the IDS alerts with the labels of the samples.
- For each sample, we get

		What the IDS says: Normal	What the IDS says: Malicious
What the label says: Normal		True Negative (TN)	False Positive (FP)
What the label says: Malicious		False Negative (FN)	True Positive (TP)

Confusion Matrix

- Such a matrix is called confusion matrix
- Typically, you write the number of samples in the matrix
- For example, for a labeled dataset with 20 malicious samples and 80 normal samples:

		What the IDS says	
		Normal	**Malicious**
What the label says	**Normal**	**70 (TN)**	**10 (FP)**
	Malicious	**5 (FN)**	**15 (TP)**

Tuning

- Ideally, False Positives and False Negatives are very close to 0
  - FP = Expensive: Alerts have to be checked.
  - FN = Security risk: Unnoticed intrusions!

- When designing an IDS (or writing rules for an IDS), a compromise has to be found
- Stupid example: Alert on all ICMP packets
  - TP = high, FN = low : All ICMP-based attacks are detected
  - But: FP = high, too.

- Acceptable FP and FN depend on your security policy

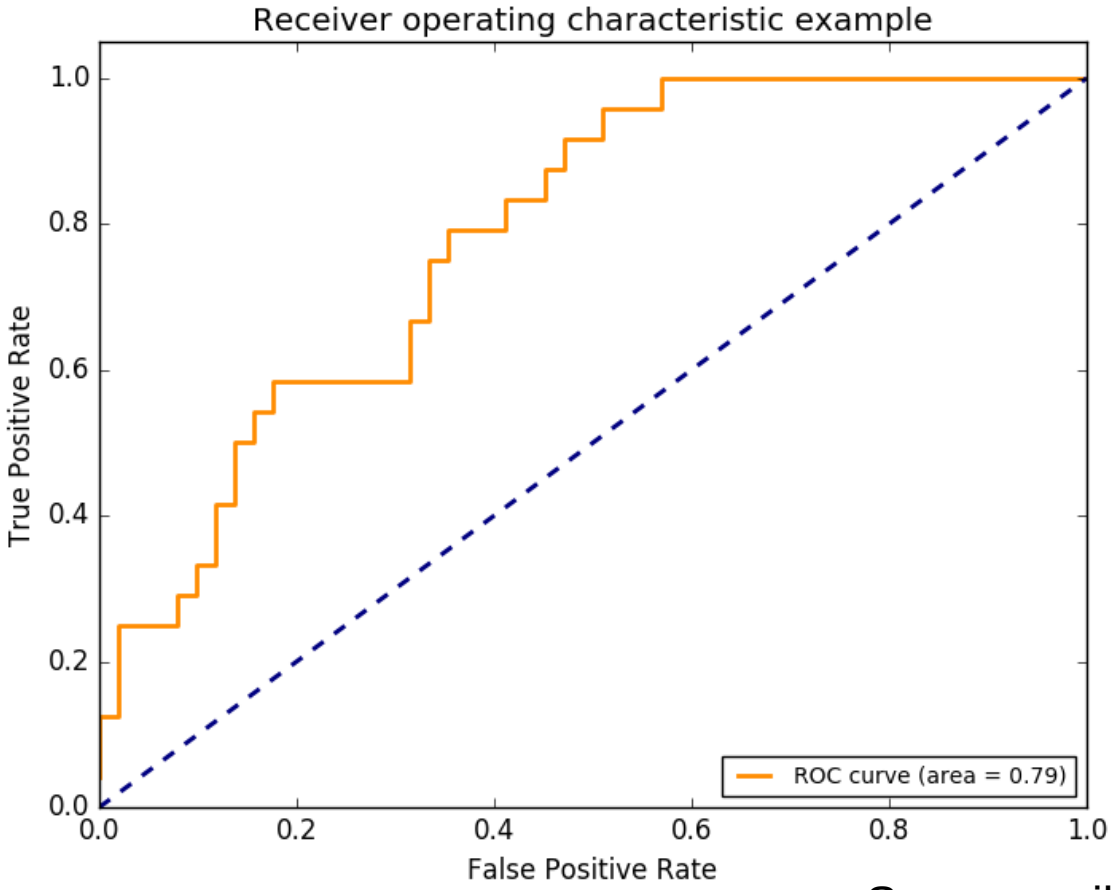
Tuning (2)

- IDS rules using thresholds have to be adapted to the target system
- Example:

  Rule: “Raise alert if UDP traffic is higher than 5 Gbps”
- The threshold is chosen based on experience. Some IDS try to “learn” the threshold automatically by observing.

Receiver Operator Characteristic (ROC)

- The dependency of the IDS performance on a specific detection parameter (e.g. threshold) can be visualized by the ROC curve



Source: scikit-learn.org