Quantitative Security

Colorado State University
Yashwant K Malaiya

CS 559

L6: Probability & Intrusion Detection

CSU Cybersecurity Center
Computer Science Dep

About this Course

CS 559 is a research-oriented course.

- 200-level classes: little student content
- 400-level: 5% student presentations/discussions
- 530: 10-15% student presentations/discussions
- 559: 25-40% student presentations/discussions

Quick Project Presentations

- Presentations coming Tuesday, Thursday
  - MS Teams
- 5 min presentations, max 7 slides
  - Submit slides 48-hours in advance on Canvas Discussions
  - Everyone should preview upcoming presentations
  - Schedule will be posted today
- 1-2 minutes discussions
- Same topic: All presents should
  - Exchange plans/documents
  - collaborate to minimize overlap.

Quantitative Security

Colorado State University

Yashwant K Malaiya

CS 559

Probabilistic Perspective

CSU Cybersecurity Center
Computer Science Dep

Conditional Probability

- Conditional probability

$$P\{A \mid B\} = \frac{P\{A \cap B\}}{P\{B\}} \; for P\{B\} > 0$$

P{A|B} is the probability of A, given we know B has happened.

- If A and B are independent, P{A|B}= P{A}. Then

$$P\{A \cap B\} = P\{A\}P\{B\}$$

- Example: A toss of a coin is independent of the outcome of the previous toss.

Conditional Probability

- If A can be divided into disjoint A_i, i=1,..,n, then

$$P\{B\} = \sum_i P\{B \mid A_i\} P\{A_i\}.$$

- Example: A chip is made by two factories A and B. One percent of chips from A and 0.5% from B are found defective. A produces 90% of the chips. What is the probability a randomly encountered chip will be defective?
- P{a chip is defective} = (1/100)x0.9 + (0.5/100)x0.1

  =0.0095 i.e. 0.95%

Bayes' Rule

- Conditional probability

$$P\{A \mid B\} = \frac{P\{A \cap B\}}{P\{B\}} \; for P\{B\} > 0$$

P{A|B} is the probability of A, given we know B has happened.

- Bayes' Rule

$$P\{A \mid B\} = \frac{P\{B \mid A\}P\{A\}}{P\{B\}} \; for\ P\{B\} > 0$$

- Example: A drug test produces 99% true positive and 99% true negative results. 0.5% are drug users. If a person tests positive, what is the probability he is a drug user?

$$P\{DU \mid P\} = \frac{P\{P \mid DU\}P\{DU\}}{P\{P \mid DU\}P\{DU\} + P\{P \mid nDU)P\{nDU\}}$$

$= \quad 33.3\%$

Confusion Matrix

	Disease +	Disease -
Test +ve	TP	FP
Test –ve	FN	TN

Evaluating a classification approach

- Precision = TP/(TP+FP) PPV positive predictive value
  - If the result is positive, what is the prob it is true?
- Several other measures used.
  - Ex: TP= 100, FP = 10, FN = 5, TN = 50
  - Precision = 100/(100+10) = 0.901

Example: Intrusion Detection

- If an ID scheme is more sensitive, it will increase false positive rates.
- Ex Car alarm

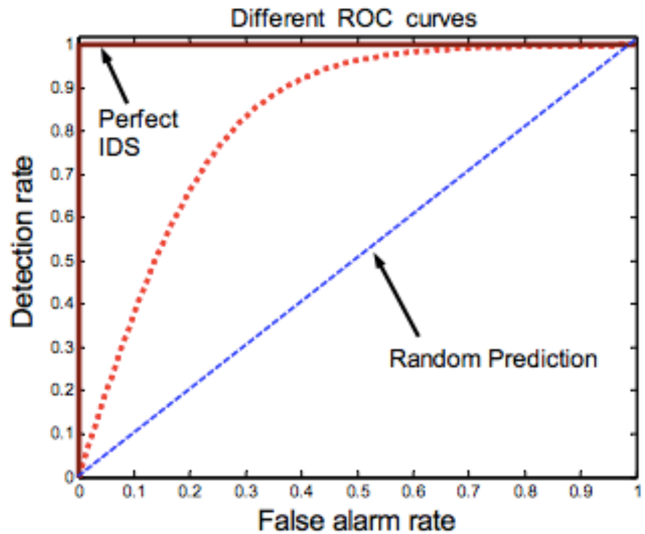


Figure 2-5. ROC Curves for different intrusion detection techniques

- True Positive rate (sensitivity) vs False Positive Rate
- Area under the ROC curve is a good measure of the ID scheme.

Intrusion Detection A Survey, Lazarevic, Kumar, Srivastava, 2008

Random Variables

- A random variable (r.v.) may take a specific random value at a time. For example
  - X is a random variable that is the height of a randomly chosen student
  - x is one specific value (say 5’9”)
- A random variable is defined by its density function.
- A r.v. can be continuous or discrete

		continuous	*discrete*
Density function	$f(x)dx$	$P\{x \le X \le x+dx\}$	$p(x_i)$
“Cumulative distribution function” (cdf)	$F(x)$	$\int_{x\min}^{x} f(x)dx$	$\sum_{i=i\min}^{i\max} p(x_i)$
Expected value (mean)	$E(X)$	$\int_{x\min}^{x\max} x f(x)dx$	$\sum_{i=i\min}^{i\max} x_i p(x_i)$

Distributions, Binomial Dist.

- Note that $$\int_{x\min}^{x\max} f(x)dx = 1 \quad \sum_{i\min}^{i\max} p(x_i) = 1$$
- Major distributions:
  - Discrete: Bionomial, Poisson
  - Continuous: Uniform, Gaussian, exponential
- **Binomial distribution**: outcome is either success or failure
  - Prob. of *r* successes in *n* trials, prob. of one success being *p*

$$f(r) = \binom{n}{r} p^r (1-p)^{n-r} \quad for \quad r = 0, \ldots, n$$

incidentally $$\binom{n}{r} = {}^nC_r = \frac{n!}{r!(n-r)!}$$

Distributions: Poisson

- **Poisson**: also a discrete distribution, λ is a parameter.

$$f(x) = \frac{\lambda^{x} e^{-\lambda}}{x!}$$

- Example: μ = occurrence rate of something.
  - Probability of r occurrences in time t is given by

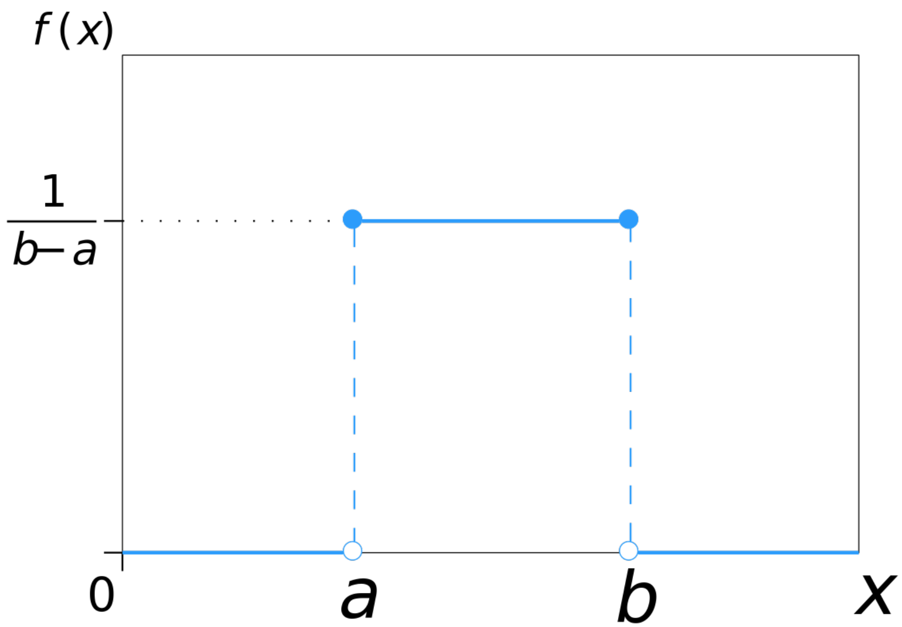
$$f(r) = \frac{(\mu t)^{r} e^{-\mu t}}{r!}$$

Often applied to fault arrivals in a system

Distributions: Uniform

- Uniform distribution:

$$f(x) = \begin{cases} 0, & x < a \\ \boxed{\dfrac{1}{b-a}}, & a \le x \le b \\ 0, & x > b \end{cases}$$



Distributions: Gaussian 1809 AD

- Continuous. Also termed Normal (called Laplacian in France! 1774 AD)

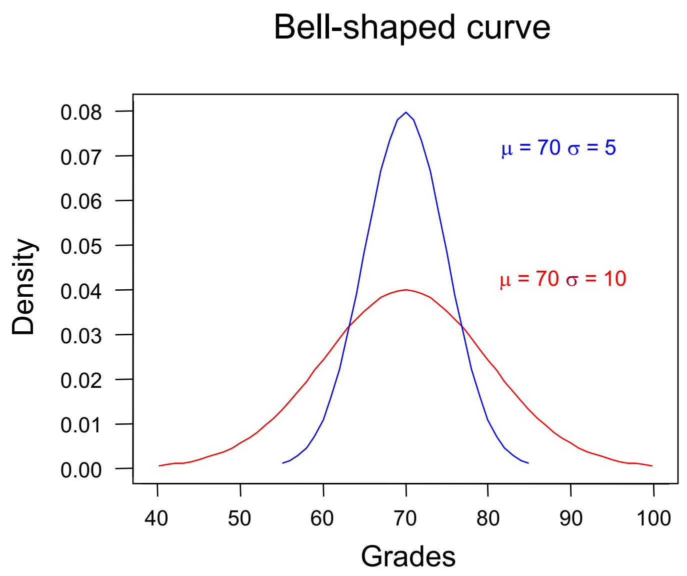
Laplace discovered it before Gauss in 1774 AD!

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

$$-\infty \le x \le +\infty$$

σ : standard deviation which is ($\sqrt{\text{variance}}$)

μ : mean



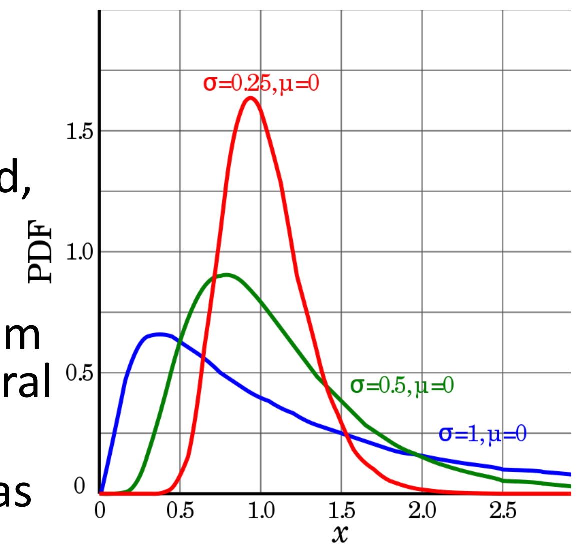
Normal distribution (2)

- Tables for normal distribution are available, often in terms of standardized variable $z=(x-\mu)/\sigma$.
- $(\mu-\sigma, \mu+\sigma)$ includes 68.3% of the area under the curve.
- $(\mu-3\sigma, \mu+3\sigma)$ includes 99.7% of the area under the curve.
- Central Limit Theorem: Sum of a large number of independent random variables tends to have a normal distribution.

The reason why normal distribution is applicable in many cases

Lognormal Distribution

- Lognormal distribution is a continuous distribution of a random variable whose logarithm is normally distributed.
  - If the random variable X is log-normally distributed, then Y = ln(X) has a normal distribution
  - A log-normal process is the realization of the multiplicative product of many independent random variables, each of which is positive. (From the central limit theorem)
  - Can't generate a zero or negative amount, but it has a tail to the right that allows for the possibility of extremely large outcomes. Often a realistic representation of the probability of various amounts of loss.
  - Widely applicable in social/technological/biological systems: file sizes, network traffic, length of Internet posts.
  - Formulas, properties: see literature.



$0 \leq X \leq \infty$

Distributions in Excel

Most common distributions are provided.

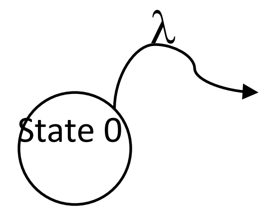
- Ex: LOGNORM.DIST(x, mean, standard_dev, cumulative)
  - X value at which you want to evaluate the log-normal function.
  - mean The arithmetic mean of ln(x).
  - standard_dev The standard deviation of ln(x).
  - Cumulative - A logical argument which denotes the type of distribution to be used:
    - TRUE = Cumulative Normal Distribution Function
    - FALSE = Normal Probability Density Function
- LOGNORM.INV(probability, mean, standard_dev)
    - Probability - The value at which you want to evaluate the inverse function.
    - Mean- The arithmetic mean of ln(x).
    - standard_dev- The standard deviation of ln(x).
- Errors: $x \leq 0$, standard_dev ≤ 0, probability ≤ 0 or ≥ 1;

Exponential & Weibull Dist.

Exponential Distribution: is a continuous distribution.

– Density function

$f(t) = \lambda e^{-\lambda t} \quad 0<t\leq\infty$

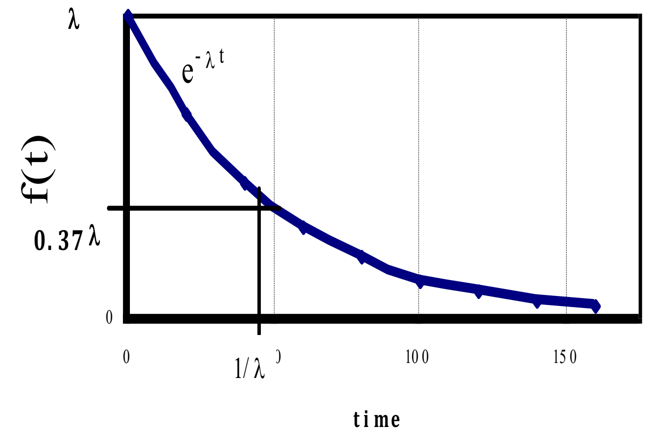


Example:

- λ: exit or failure rate.
- Pr{exit the good state during (t, t+dt)}

  $= e^{-\lambda t} \lambda\, dt$

- The time T spent in good state has an exponential distribution



- Weibull Distribution: is a 2-parameter generalization of exponential distribution. Used when better fit is needed, but is more complex.

Variance & Covariance

- Variance: a measure of spread
  - $Var\{X\} = E[X-\mu_x]^2$
  - Standard deviation = $(Var\{x\})^{1/2}$
  - σ = standard deviation (usually for normal dist)
- Covariance: a measure of statistical dependence
  - $Cov\{X,Y\} = E[(X-\mu_x)(Y-\mu_y)]$
  - Correlation coefficient: normalized

    $\rho_{xy} = Cov\{X,Y\}/ \sigma_x \sigma_y$

  Note that $0<|\rho_{xy}|<1$

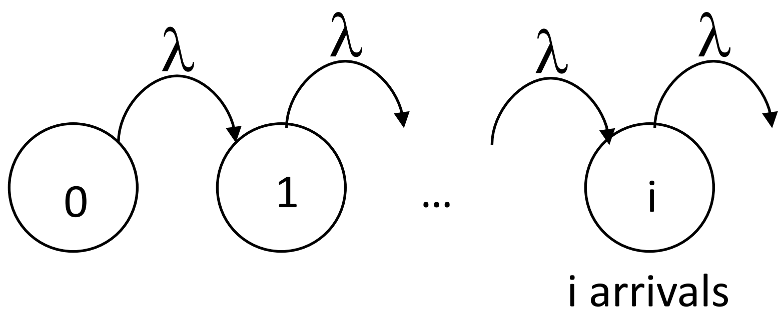
Stochastic Processes

- Stochastic process: that takes random values at different times.
  - Can be continuous time or discrete time
- Markov process: discrete-state, continuous time process. Transition probability from state i to state j depends only on state i (It is memory-less)
- Markov chain: discrete-state, discrete time process.
- Poisson process: is a Markov counting process N(t), $t \geq 0$, such that N(t) is the number of arrivals up to time t.

Poisson Process: properties

- Poisson process: A Markov counting process N(t), $t \geq 0$, N(t) is the number of arrivals up to time t.
- Properties of a Poisson process:
  - N(0) = 0
  - P{an arrival in time Δt} = $\lambda \Delta t$
  - No simultaneous arrivals
- We will next see an important example. Assuming that arrivals are occurring at rate λ, we will calculate probability of n arrivals in time t.

- A process is in state I, if I arrivals have occurred.
- $P_i(t)$ is the probability the process is in state i.

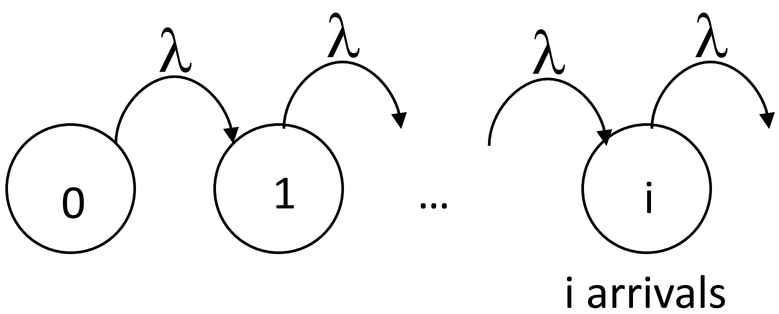


- In state i, probability is flowing in from state i-1, and is flowing out to state i+1, in both cases governed by the rate λ. Thus

$$\frac{dP_i(t)}{dt} = -\lambda P_i(t) + \lambda P_{i-1}(t) \quad n = 0,1,..$$

We'll solve it first for $P_0(t)$, then for $P_1(t)$, then …

Poisson process: Solution for $P_0(t)$



$$P_0 = P\{process\ \ in\ state\ \ 0\}$$

$$P_0(t+\Delta t) = P_0(t)[1-\lambda\Delta t]$$

$$\frac{P_0(t+\Delta t) - P_0(t)}{\Delta t} = -\lambda P_0(t)$$

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t)$$

$$Solution:$$

$$\ln(P_0(t)) = -\lambda t + C$$

$$P_0(t) = C_2 e^{-\lambda\, t}$$

$$Since\ P_0(0) = 1,\ C_2 = 1,$$

$$\boxed{P_0(t) = e^{-\lambda\, t}}$$

Poisson Process: General solution

We need to solve
$$\frac{dP_i(t)}{dt} = -\lambda P_i(t) + \lambda P_{i-1}(t) \quad n = 0,1,..$$

Using the expression for $P_0(t)$, we can solve it for $P_1(t)$.

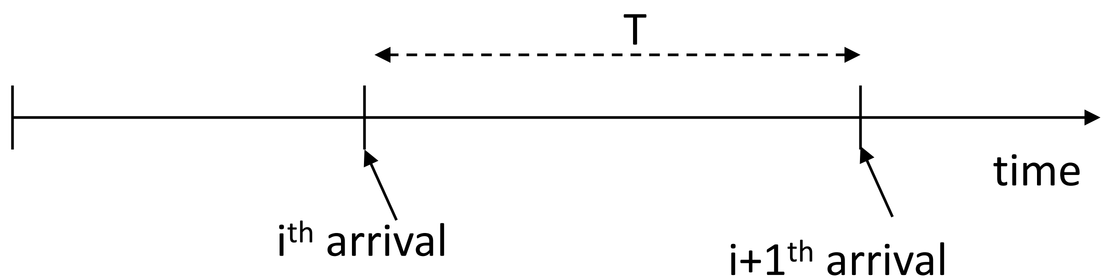
$$Solving\ \ recursively, we\ \ get$$

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda\ t} \quad n = 0,1,..$$

Which we know is Poisson distribution!

Poisson Process: Time between Two Events

Here we'll show that the time to next arrival is exponentially distributed.



$$P\{t_{i+1} > t\} = P\{no\ arrival\ in\ (t_i, t_i + t)\} = e^{-\lambda t}$$

Thus the cumulative distribution function (cdf) is given by

$$F(t) = P\{0 \le T \le t\} = 1 - e^{-\lambda t}$$

Since the density function is derivative of cdf,

differentiating both sides, we get

$$f(t) = \lambda e^{-\lambda t}$$

Exponential distribution

Quantitative Cyber-Security

Colorado State University

Yashwant K Malaiya

Fall 2020

Intrusion Detection

CSU CyberCenter
***Course Funding Program* – 2019**

Cyber-security/cybersecurity/Cyber security?

Intrusion Detection

- Intrusion: Unauthorized act of bypassing the security mechanisms of a system.
- Intrusion Detection System (IDS): A software/hardware system that gathers and analyzes information to identify possible intrusions
  - from various areas within a computer (Host-based) **HIDS**
    - Monitors the characteristics of a single host for suspicious activity
  - Traffic on a a network (network based) **NIDS**
    - Monitors network traffic and analyzes network, transport, and application protocols to identify suspicious activity
  - Hybrid
- IDS components:
  - "Sensors" - collect data
  - Analyzers - determine if intrusion has occurred
  - User interface - view output or control system behavior

Intrusion Detection

IDS Detection Approaches

Two approaches

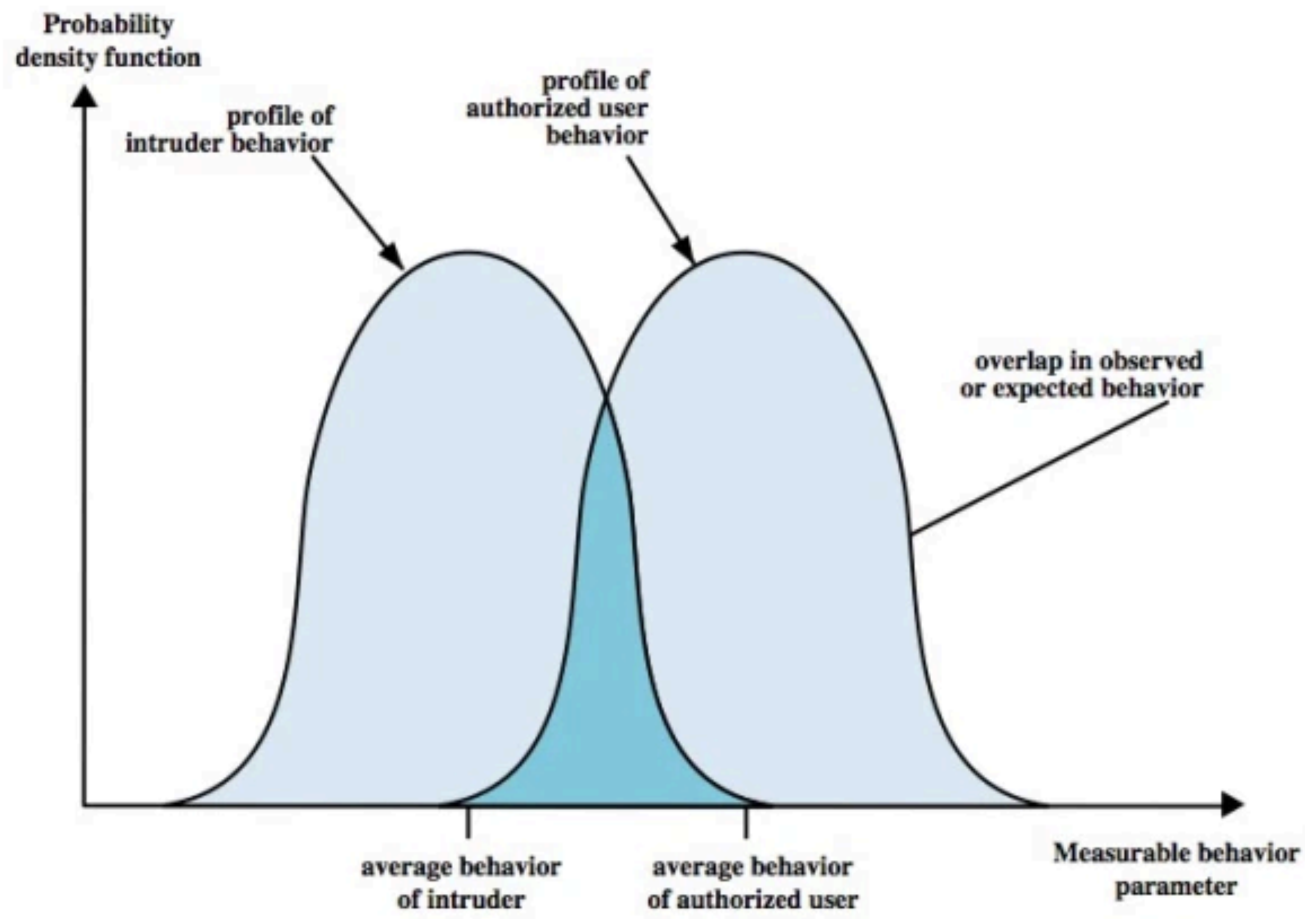
- Anomaly detection: Is this the normal behavior?
  - Collection of data about the behavior of legitimate users
  - Does the current behavior resemble that of a legitimate user?
- Signature based detection: Does it match known bad behavior?
  - Match a large collection of known patterns of malicious data against data on a system or in transit over a network
- Rule-based heuristic
  - Rules that identify suspicious behavior

Stallings and Brown, 4th ed.

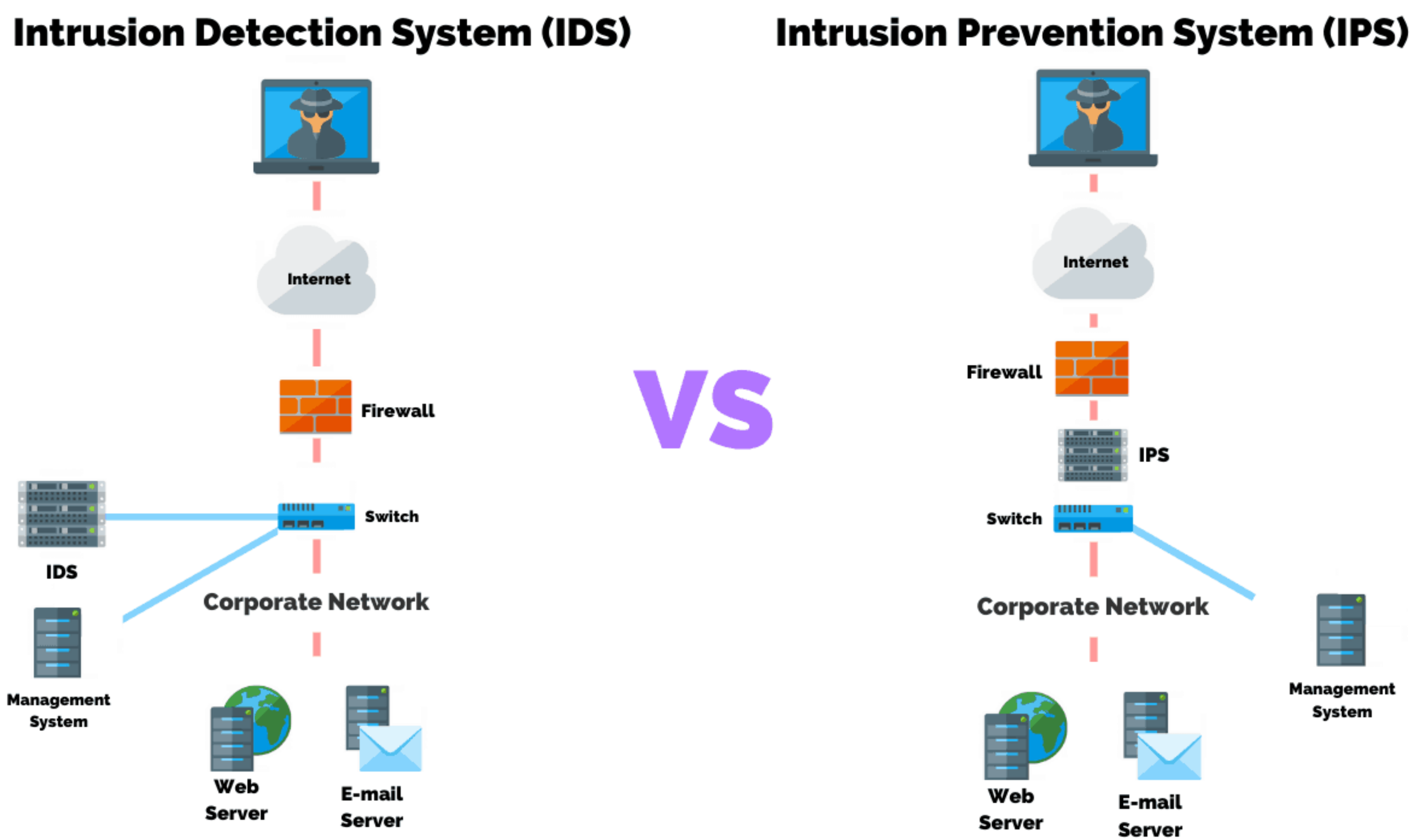
Colorado State University

Intruder vs normal behavior

No clear diving line between intruder vs authorized user activity



IDS vs IPS



https://purplesec.us/intrusion-detection-vs-intrusion-prevention-systems/

Details

Host-Based Intrusion Detection (HIDS)

- a specialized layer of security software
- either anomaly or signature and heuristic approaches
- Monitors activity to detect suspicious behavior
  - to detect intrusions, log suspicious events, and send alerts
  - Monitors system calls, DLL activity
  - Can detect both external and internal intrusions

NIDS: information logged by a NIDS sensor includes

- Timestamp
- Connection or session ID
- Event or alert type
- Rating
- Network, transport, and application layer protocols
- Source and destination IP addresses
- Source and destination TCP or UDP ports, or ICMP types and codes
- Number of bytes transmitted over the connection
- Decoded payload data, such as application requests and responses
- State-related information

Intrusion Detection Techniques

Signature Detection can effective for	Anomaly detection can be effective for
• Application layer reconnaissance and attacks • Transport layer reconnaissance and attacks • Network layer reconnaissance and attacks • Unexpected application services • Policy violations	• Denial-of-service (DoS) attacks • Scanning • Worms

Colorado State University

IDS Examples

- Antivirus: looks for signatures of known threats
- SNORT: a multi-mode packet analysis tool
  - Sniffer, Packet Logger, Forensic Data Analysis too, Network Intrusion Detection System
  - Rules form “signatures”
    - Modular detection elements are combined to form these signatures
    - Wide range of detection capabilities
      - Stealth scans, OS fingerprinting, buffer overflows, back doors, CGI exploits, etc.
    - Rules system is very flexible, and creation of new rules is relatively simple.
    - bad-traffic.rules, exploit.rules, scan.rules, smtp.rules, smtp.rules, backdoor.rules shellcode.rules ….

Example study

Performance comparison of intrusion detection systems and application of machine learning to Snort system

- Shah and Isaac, 2017
- Two open source IDS Snort and Suricata compared, with specific algorithms
- Normal and malicious traffic, different protocols
- Positive = TP+FN, Negative = FP+TN
- FPR = FP/(FP+TN), FNR = FN/(FN+TP)

Malicious traffic accuracy (%) measurements at 10 Gbps during 12 h.

Malicious Traffic	First 4 h			
	Snort FPR	Snort FNR	Suricata FPR	Suricata FNR
SSH	8.0	0.0	7.0	0.0
DoS/DDoS	3.0	2.0	10.0	0.0
FTP	12.0	0.0	11.0	0.0
HTTP	5.0	0.0	8.0	0.0
ICMP	20.0	0.0	22.0	0.0
ARP	7.0	0.0	10.0	12.0
Scan	1.0	4.0	5.0	3.0
Total	56.0	6.0	73.0	15.0

Colorado State University