Quantitative Cyber-Security

Colorado State University
Yashwant K Malaiya

CS559

Probability, Intrusion Detection



CSU Cybersecurity Center
Computer Science Dept

Probability: FAQ

Probability: According to some "one of the scariest things in mathematics".

Note that the algebra involving the probabilities is often simple.

- It is **going from a verbal description to an algebraic representation** of the problem, where the challenge often lies.

You should be clear about:

- Are the events A and B disjoint?
- Are the events A and B statistically independent?
- Is the probability we are considering Pr{A|B} or is it {Pr{A}?

Confusion Matrix

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

- Sensitivity (Recall) = TP/(TP+FN) also TPR true pos rate
  - If the person has the disease, what is the prob test is positive?
  - (TP+FN) have the disease, TP have have a positive test.
  - Hence Sensitivity = TP/(TP+FN)
- Specificity = TN/(FP+TN) also TNR true neg rate
  - If the person does not have the disease, what is the prob test is indeed negative?
  - (FP+TN) do not have the have disease, TN test negative.
  - Hence Specificity = TN/(FP+TN
- Precision = TP/(TP+FP) PPV positive predictive value
  - If the result is positive, what is the prob it is true?
  - (TP+FP) have the positive result, of them TP have the disease
  - Hence precision = TP/(TP+FP)
- Accuracy = (TP + TN) / (TP + TN + FP + FN).
  - Fraction of results that are correct
  - Misleading when data is unbalanced
- There are several other metrics, For example
  - F1 score is *harmonic mean* of precision and recall
  - 2TB/(2TP + FP + FN)

RT-PCR s	Disease +	Disease -
Test +ve	580	21
Test –ve	308	105

HM = n / Σ(1/xi)

Bayes' Rule

Example: A drug test produces 99% true positive and 98% true negative results. 0.5% are drug users. If a person tests positive, what is the probability he is a drug user?

$$P\{DU|P\} = \frac{P\{P|DU\}P\{DU\}}{P\{P|DU\}P\{DU\}+P\{P|nDU\}P\{nDU\}}$$

$$= \frac{0.99\times0.005}{0.99\times0.005+0.02\times0.995} = 0.1991$$

$$= 19.91\%\%$$

- Person is a drug user and test result is positive: True Positive
- P{nDU} = 99.5%,
- P{P|nDU} = 1-0.98 = 0.02

Bayes' rule explanation

Example: A drug test produces 99% true positive and 98% true negative results. 0.5% are drug users. If a person tests positive, what is the probability he is a drug user?

- $$P\{DU|P\} = \frac{P\{P|DU\}P\{DU\}}{P\{P|DU\}P\{DU\}+P\{P|nDU\}P\{nDU\}}$$

- 99% true positive means 99% of the actual drug users will test positive.
  - TPR = TP/actual drug user = TP/(TP+FN)
- 98% true negative means 98% of the non-drug users will test negative.
  - TNR = TN/actual non-user = TN/(TN+FP)
- We know P{P/DU} = 0.99. But we need to find P{DU/P}. Thus you need **Bayes' rule**.
- Numerator is 0.99x(0.5/100)
- The second term of the denominator is (1-0.98)x(99.5/100)
- **Find and view Short videos**: Binomial dist., Bayes' theorem

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

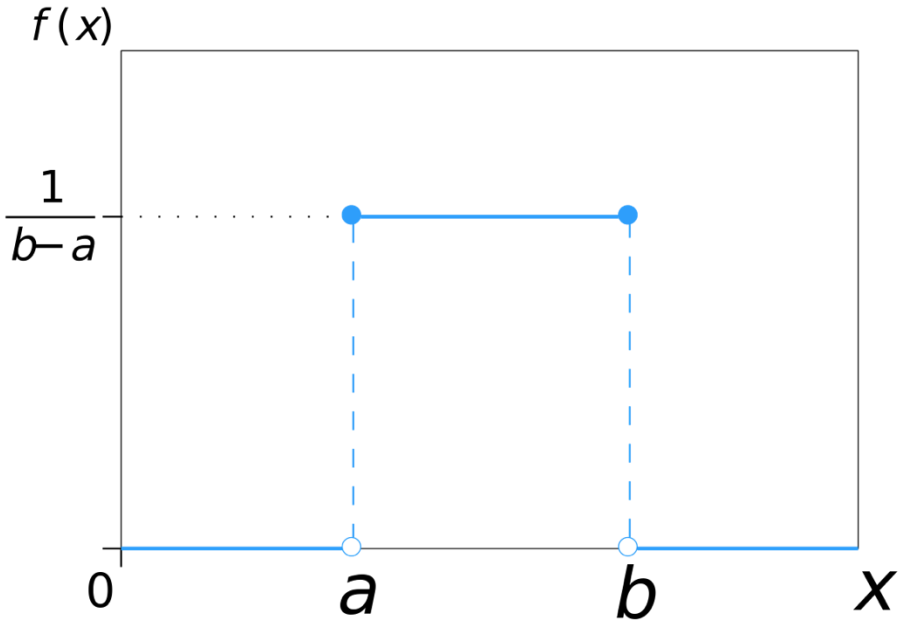
$$f(r) = \frac{(\mu t)^{r} e^{-\mu t}}{r!}$$

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)

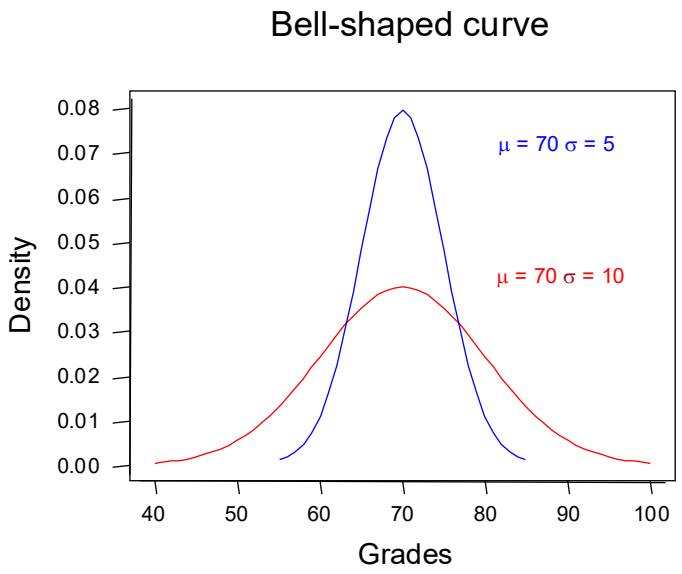
Laplace discovered it before Gauss in 1774 AD!

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

$$-\infty \leq x \leq +\infty$$

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

μ : mean



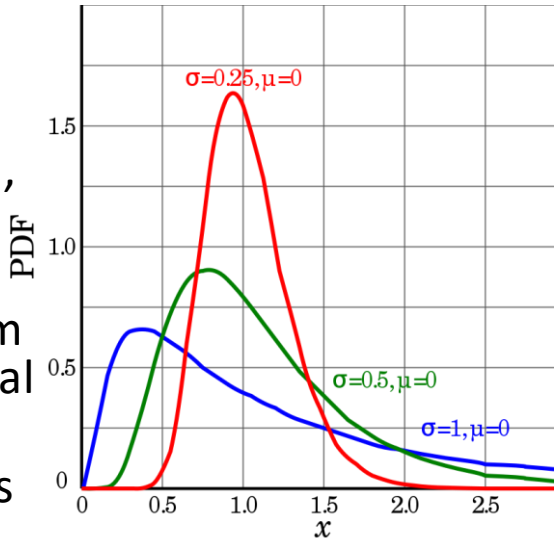
Normal distribution (2)

- 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.

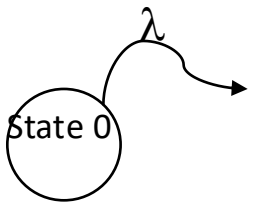
- Ex: LOGNORM.DIST(x, mean, standard_dev, cumulative)
  - X value at which you want to evaluate the log-normal function.
  - mean The arithmetic mean of ln(x).
  - standard_dev The standard deviation of ln(x).
  - Cumulative - A logical argument which denotes the type of distribution to be used:
    - 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

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

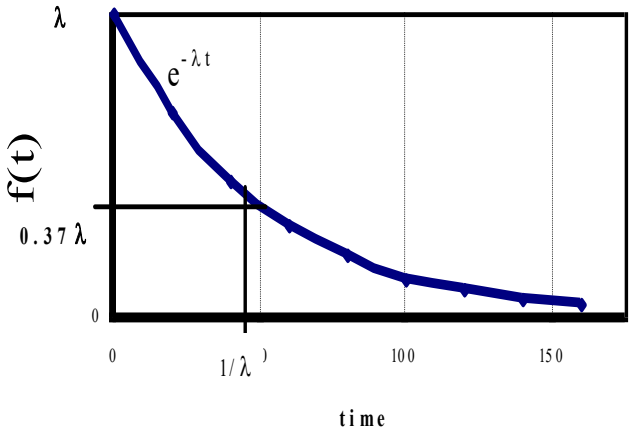


Example:

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

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

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

Stochastic Processes

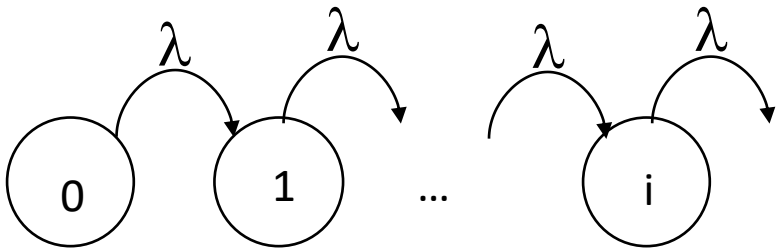
- Stochastic process: that takes random values at different times.
  - Can be continuous time or discrete time
- Markov process: discrete-state, continuous time process. Transition probability from state i to state j depends only on state i (It is memory-less)
- Markov chain: discrete-state, discrete 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.

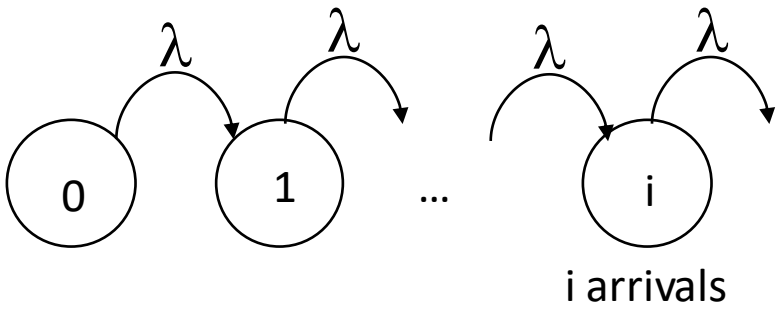


- In state i, probability is flowing in from state i-1, and is flowing out to state i+1, in both cases governed by the rate λ. Thus

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

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 $P_1(t)$.

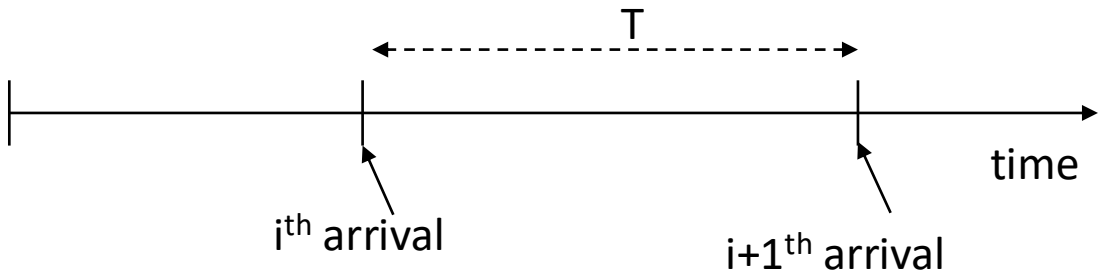
$$Solving\ recursively, we\ get$$

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

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

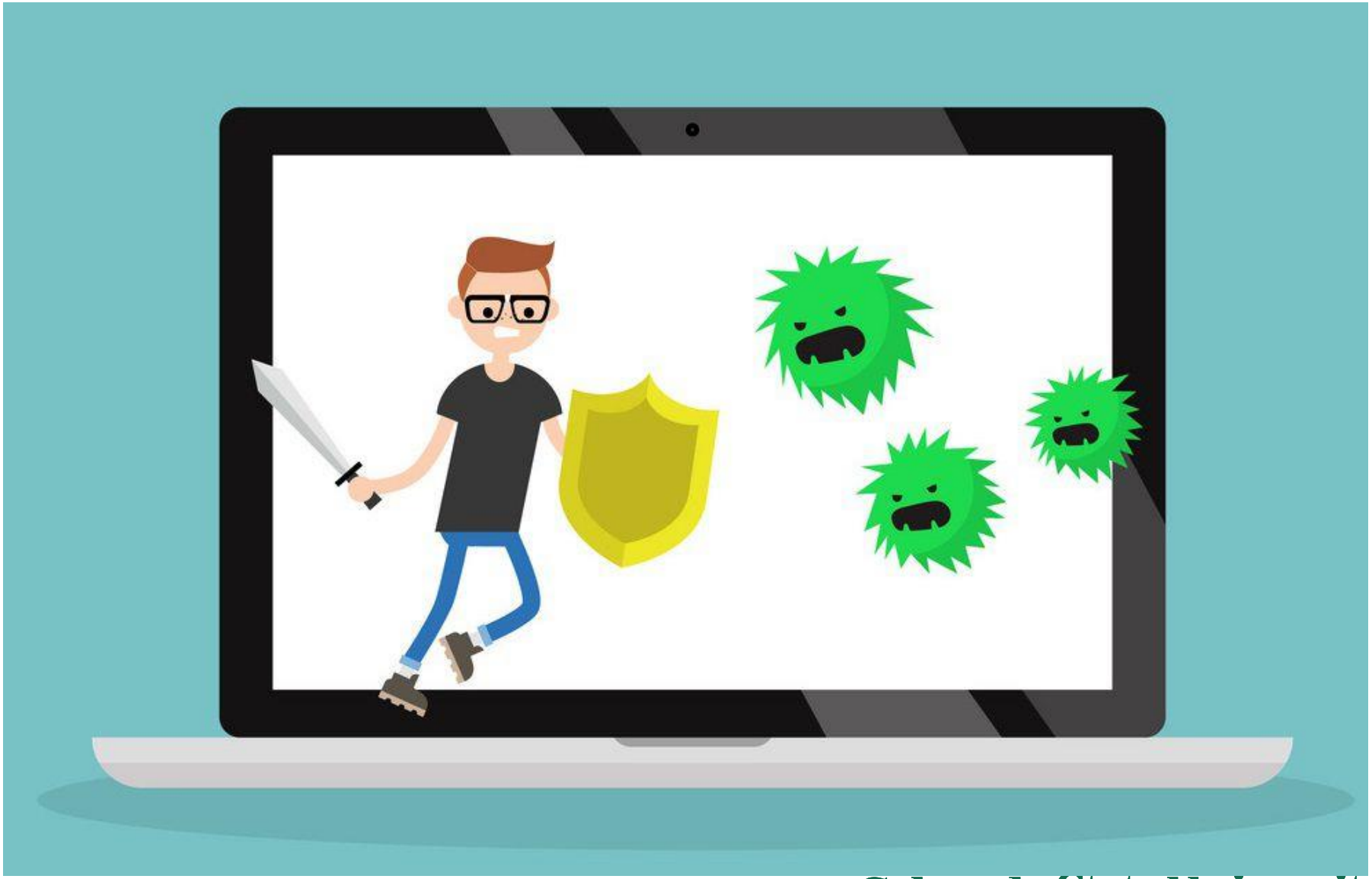
Intrusion Detection



CSU CyberCenter

Cyber-security/cybersecurity/Cyber security?

Intrusion Detection



Intrusion detection system (IDS)

- An intrusion detection system (IDS) is a security tool
  - monitors network traffic and devices for known malicious activity, suspicious activity or security policy violations.
  - alerts security administrators to known or potential threats, or to a centralized security tool.
  - Certain regulations, such as the Payment Card Industry Data Security Standard (PCI-DSS), require organizations to implement intrusion detection measures.
  - Often integrated with, or incorporated into, intrusion prevention systems (IPSs), which can detect security threats and automatically prevent them.

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

IDS Detection Approaches

Approaches

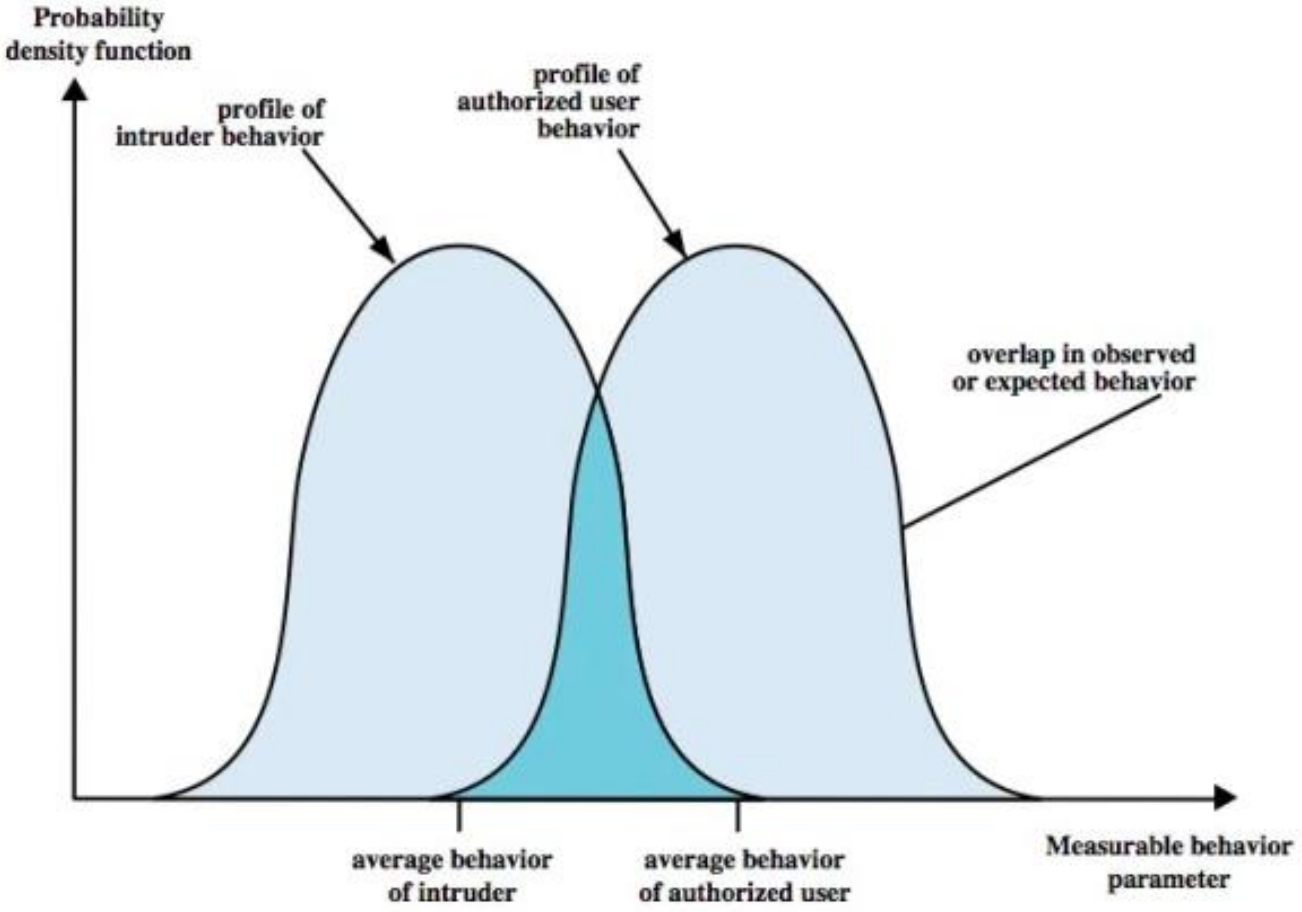
- **Signature based detection**: Does it match known bad behavior?
  - Match a large collection of known patterns of malicious data against data on a system or in transit over a network
  - Low false positives but Ineffective for a new exploit.
- **Anomaly detection**: Is this the normal behavior?
  - Collection of data about the behavior of legitimate users, use AI
  - Does the current behavior resemble that of a legitimate user?
- Reputation-based :
  - blocks traffic from IP addresses/domains associated with malicious or suspicious activity.
- Stateful protocol analysis:
  - protocol behavior, for example, a denial-of-service (DoS) attack by detecting a single IP address making many simultaneous TCP connection requests in a short period.

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

Intruder vs normal behavior

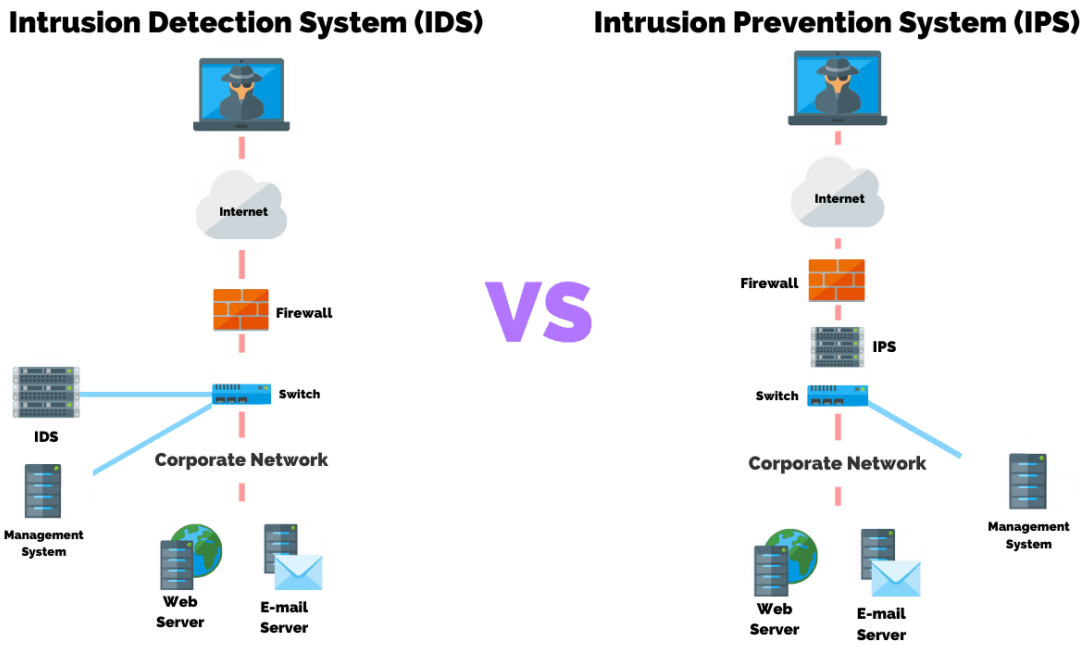
No clear diving line between intruder vs authorized user activity



IDS and IPS, Firewals

- IPS (intrusion prevention systems)
  - monitors network traffic for suspicious activity, like an IDS,
  - *intercepts* threats in real time by automatically terminating connections or triggering other security tools.
  - Because IPSs are meant to stop cyberattacks, they're placed inline, all traffic has to pass through the IPS before it can reach the rest of the network.
  - Often **IDS and IPS are combined** in a single intrusion detection and prevention system (IDPS)
- Firewalls face outside the network and act as barriers by using predefined rulesets to allow or disallow traffic.
- IDSs often sit near firewalls and help catch anything that slips past them.
- Some firewalls, especially next-generation firewalls, have built-in IDS and IPS functions.

IDS vs IPS



The IDS is designed to **detect** anomalous activity by observing the network traffic patterns. If detected, it generates an alert. The IDS is not placed directly in line with the network traffic.

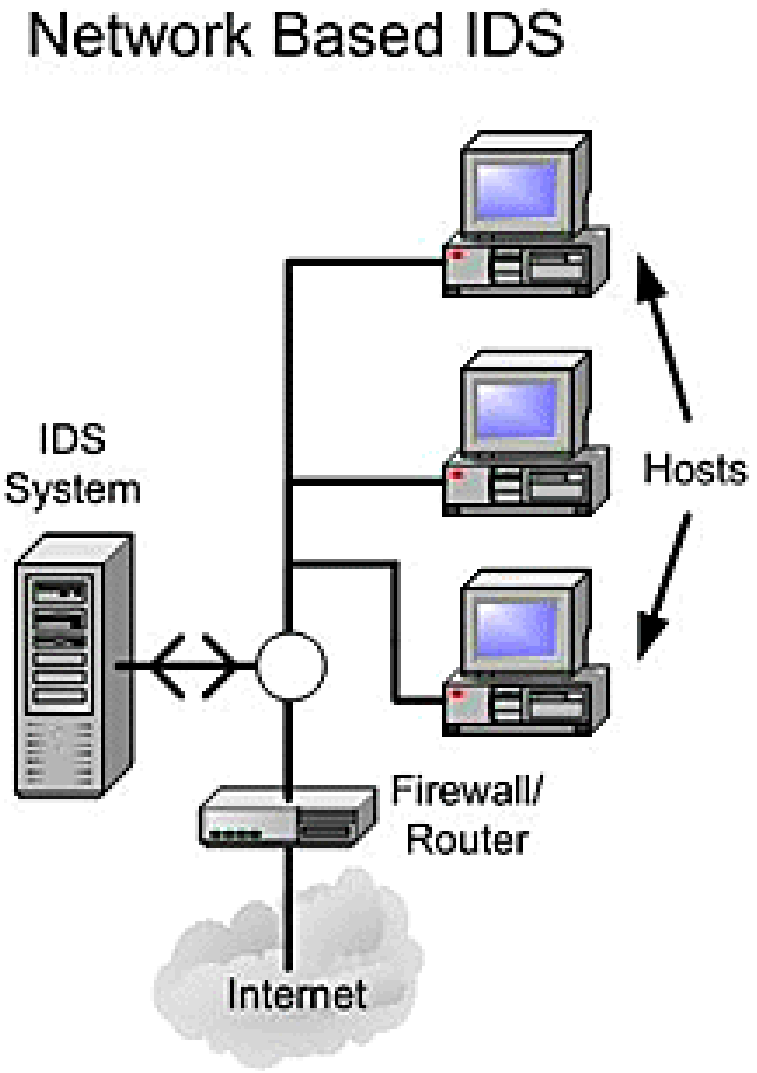
An IPS examines network traffic flows in order to **detect and prevent** vulnerability exploits. The IPS often sits directly behind the firewall.

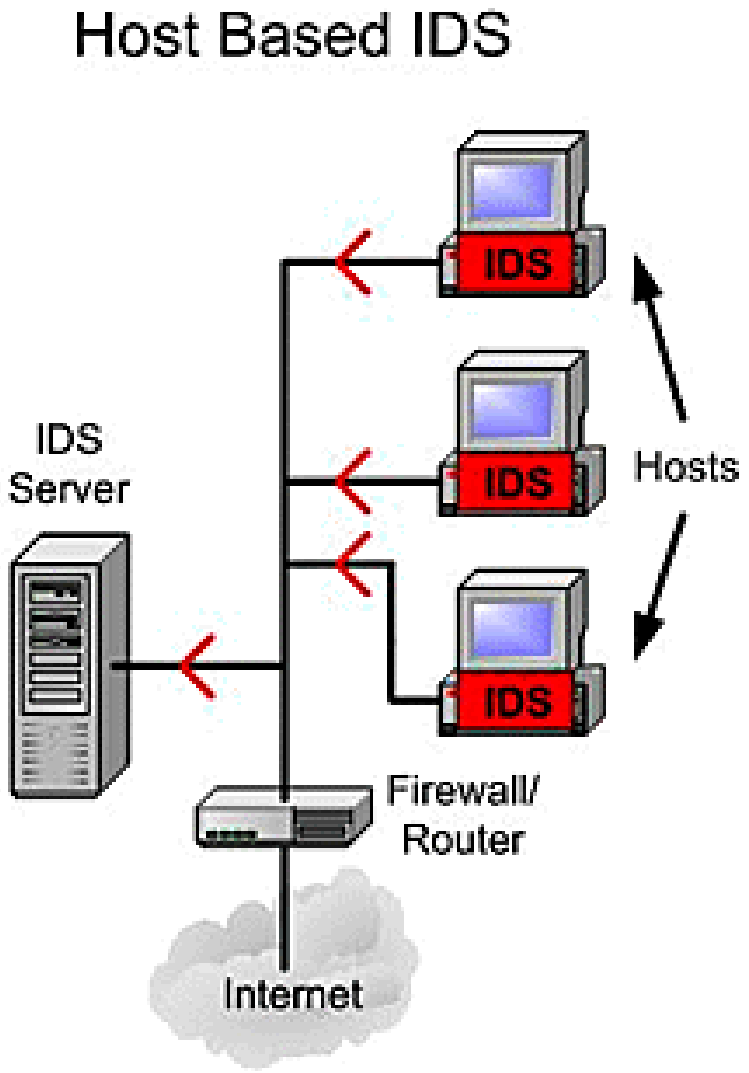
https://purplesec.us/intrusion-detection-vs-intrusion-prevention-systems/

Common Usage

- **At the Internet Edge (High IPS Use):** Inbound traffic from the raw internet to your network edge almost always passes through an active, aggressive IPS mode.
  - Malicious port scans, known automated botnets, and blunt web exploits are blocked automatically to drop noise.
- **Deep in the Internal Core (Low IPS / High IDS Mode):** Within internal core data loops, engineering networks, and highly complex application architectures, security operations teams deliberately flip systems into passive IDS monitoring mode.
  - The catastrophic business operational risk of a false positive blocking a mission-critical database sync or production industrial pipeline outweighs the desire for automated blocking. Teams prefer to receive immediate alerts, allowing humans or automated system to manage the event.

Network based vs Host based IDS





https://en.bmstu.wiki/index.php?title=HIDS_(Host-Based_Intrusion_Detection_System)&mobileaction=toggle_view_mobile

Host-Based vs Network-based Intrusion Detection

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

Network-based intrusion detection system (NIDS): Traffic information in packets 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

IDS evasion tactics

- **Distributed denial-of-service (DDoS) attacks:** taking IDSs offline by flooding them with obviously malicious traffic from multiple sources. When the IDS's resources are overwhelmed by the decoy threats, the hackers sneak in.
- **Spoofing**: faking IP addresses and DNS records to make it look like their traffic is coming from a trustworthy source.
- **Fragmentation**: splitting malware or other malicious payloads into small packets, obscuring the signature and avoiding detection. By strategically delaying packets or sending them out of order, hackers can prevent the IDS from reassembling them and noticing the attack.
- **Encryption:** using encrypted protocols to bypass an IDS if the IDS doesn't have the corresponding decryption key.
- **Operator fatigue**: generating large numbers of IDS alerts on purpose to distract the incident response team from their real activity.

https://www.ibm.com/think/topics/intrusion-detection-system

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 NIDS 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)

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

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

Industry Comparison Matrix

Intrusion Detection System	Primary Type	Average True Positive Rate (TPR)	Average False Negative Rate (FNR)		Main Operational Constraint
Snort (Cisco)	Network (NIDS)	98.50%	1.50%	Open source, widely used	Packet dropping under extreme traffic spikes.
Suricata	Network (NIDS)	97.40%	2.60%	Open source, widely used	Demands higher baseline memory and CPU resources.
Zeek	Network Behavioral	96.40%	3.60%		Requires custom script tailoring to prevent alert fatigue.
OSSEC	Host (HIDS)	98.00%	2.00%		Limited visibility over broad raw network traffic lanes.
Commercial NGFWs (Check Point, Palo Alto)	Hybrid / Integrated	99.70%	0.30%	**65% to 70%** of total enterprise market spending	High license cost and cloud ecosystem lock-in.

Based on multiple sources

Quantitative Security

Colorado State University

Yashwant K Malaiya

CS 599

Modeling and Regression



CSU CyberCenter

***Course Funding Program* – 2019**

Modeling & Regression

Model vs Reality

- Reality is complex, often too complex
- A model represents the actual behavior of interest with acceptable accuracy
  - System design & evaluation of alternatives
  - System simulation
  - Predictions
- When there are multiple models, how do you choose one?
  - Familiarity, convenience?
  - Goodness of fit?
  - Predictive capabilities?

Quantitative models

- **Derived from first principles**:
  - Arguments are actual things measured
  - Example: monthly mortgage payment

$$c = \frac{rP}{1-(1+r)^{-N}}$$

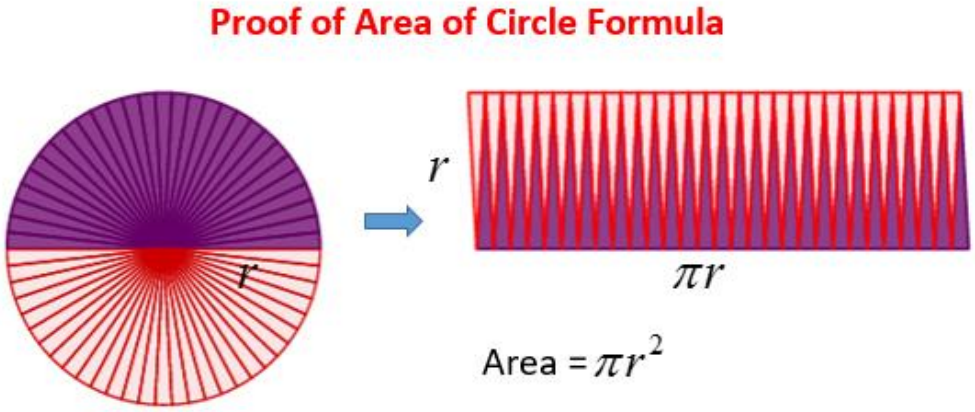
- **Empirical**
  - Arguments are just parameters
  - Example Ideal body weight in kg = a+b (height in in. - c)

    Men: a=52 kg, b= 1.9, c = 60 in.
- **Combined**
  - Empirical with interpretation of parameters
  - From first principles, adjusted to fit

41

Derived using first principles

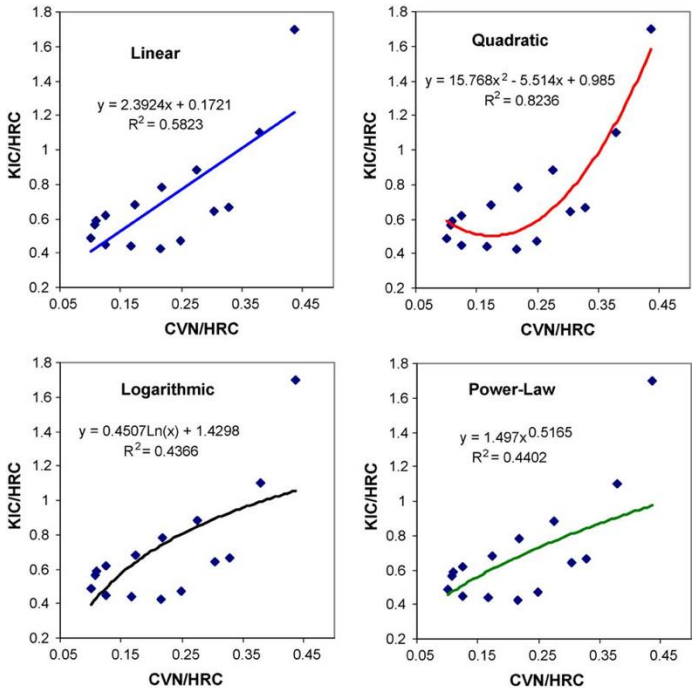
- Obtain model
  - Understand how things work
  - Use approximations
  - Derive formulas
- Validate:
  - Get real data
  - See if it fits. If not,
    - Make adjustments
    - Get alternative models
- Use the model
  - Plug in values to estimate parameters
  - Make projections



42

Empirical models

- Look at data
- See if it resembles a function
  - Linear, quadratic, logarithmic, exponential..
  - Involving 1, 2 or more parameters
- See if it fits
  - If not try something more complex
- If it fits, see if an interpretation of the parameters is possible
  - Not necessary but will be good.



43

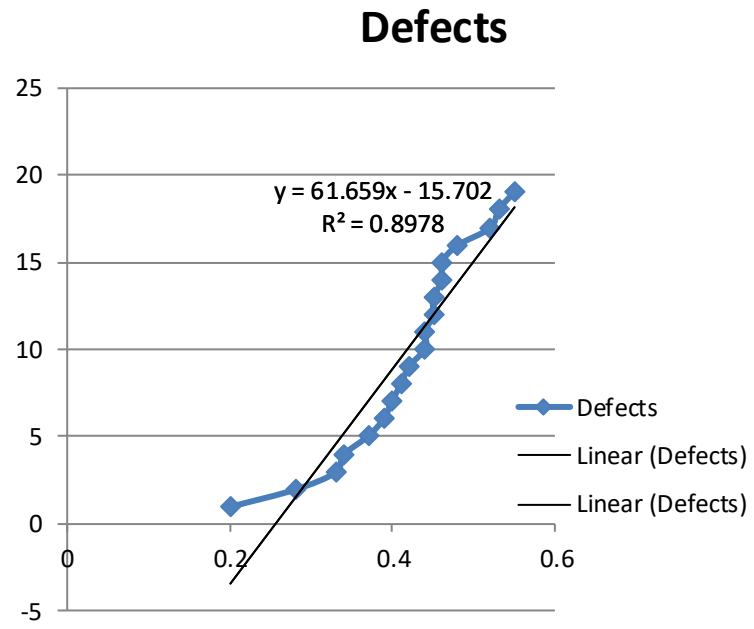
Curve fitting

- Get x-axis and y-axis numbers (x_i, y_i).
- Draw a scatter plot.
- Fit (i.e. estimate parameters):
  - Use formulas for parameters
  - linest( ) or logest( ) functions in excel
  - Select plot, rc, add trendline, select display options in excel
  - Use Solver for general fitting for excel
  - Goodness of fit measures: R^2
- Note: High goodness of fit does not imply high predictive capability

44

Example

Branch Cov	Defects
0.2	1
0.28	2
0.33	3
0.34	4
0.37	5
0.39	6
0.4	7
0.41	8
0.42	9
0.44	10
0.44	11
0.45	12
0.45	13
0.46	14
0.46	15
0.48	16
0.52	17
0.53	18
0.55	19



Open Excel sheet CurveFittingExamples.xlsx

45