Risk Assessment on Cyber-attacks: Evaluating success probability and detection accuracy

Dymphna Thomas 40267079

Manraj Kaur 40261914

***Abstract*— Given the increasing number of threats and intrusions to organizations, the use of Intrusion Detection Systems (IDS) has increased drastically. The data collected by these Intrusion Detection Systems (IDS) provides a useful pattern for risk analysis. In this paper we study the different risk parameters of an IDS system. The aim of this study is to be able to identify parameters causing success probability of an attack and provide measures to reduce the attack percentage. Risk patterns in the dataset are studied using a variety of distributions and analyzed using risk analysis methodologies. The results of the risk analysis are studied to provide key takeaways for an organization's risk treatment and mitigation approach.**

Keywords*— *risk analysis, IDS, parameters, attack, data

I. Introduction

The modern cybersecurity threat landscape demands rigorous risk assessment to precisely prioritize high risk events and defense mechanisms promptly. This paper provides a multi-layered risk analysis framework using a dataset from an IDS to quantify the intrusion risks in the network.
The aim of this paper is to identify interdependencies of the various risk parameters, for example how encryption standards would strengthen or weaken organizations' attack surface, perform cost-benefit analysis of evaluating the profitable outcome whilst managing risks, studying the attack pathways using tree analysis to put forth countermeasures and tradeoffs.

This report addresses the constant challenge of translating raw IDS data into actionable insights for risk mitigation. With the use of statistical and probabilistic methodologies, we assess and analyze the parameters influencing cybersecurity risks, such as failed login attempts, session durations, protocol types, and encryption standards. The objective of this project is to identify high-risk factors with IDS datasets through rigorous analysis and to provide a framework for organizations to prioritize security enhancements.

Our methodology is a combination of diverse risk analysis techniques, including parametric distributions (Poisson, Binomial, Normal, and Weibull) to model discrete and continuous risk events, Bayesian analysis to quantify conditional probabilities of attacks, and decision-support tools like Fault Tree Analysis (FTA) and Sensitivity Analysis to evaluate mitigation strategies. By integrating these approaches, we bridge theoretical risk models and cybersecurity decision-making.

Our project highlights both statistical risk indicators, such as the correlation between weak encryption protocols and attack success, and data-driven assessments to transform security postures into active resilient architectures. By doing this, we provide evidence-based methodology for risk prioritization and mitigation.

II. Dataset

The dataset we used is the Cybersecurity Intrusion Detection Dataset, publicly available on Kaggle [1]. It consists of 9,537 labelled network sessions with binary attack indicators (attack_detected: 0 for normal, 1 for attack). It includes the following features:
Numerical Parameters:

- failed_logins, which is the discrete count of failed authentication attempts per session.
- session_duration is the continuous duration of sessions (in seconds).
- ip_reputation_score is the score reflecting the historical trustworthiness of IP addresses (0–1 scale).
- network_packet_size is the size of transmitted data packets (in bytes).

Categorical Parameters:

- protocol_type is the network protocols like TCP, UDP, and ICMP.
- encryption_used indicates the encryption standards, like AES, DES.
- browser_type is the browser identifiers (Chrome, Firefox).
- unusual_access_time is the Binary flag (0/1) for sessions during non-standard hours.

A. Dataset Utility and Limitations

The dataset enabled robust risk modelling but introduced constraints:

- Assumptions made were:
  - Session independence (potential real-world attack correlations ignored).
  - Linearity in variable relationships (non-linear patterns not addressed).
- Limitations:
  - Right-censoring: 55.29% of sessions lacked attacks, complicating survival analysis.
  - Non-normal distributions: session_duration exhibited right skewness (Shapiro-Wilk $p = 0.000$).

- Class imbalance: Attack vs. non-attack ratios (44.71% vs. 55.29%) may bias probabilistic inferences.

This dataset provides a foundational resource for classical risk analysis in intrusion detection, aligning with methodologies such as Poisson modelling of brute-force attempts [2] and Bayesian assessment of encryption risks [3]. The future work should be in its combination with machine learning for addressing non-linear relationships and to validate findings across different types of network environments.

III. Problem Description

A. System Overview

We analysed the network infrastructure protected by an Intrusion Detection System (IDS). Its key components include:

1. Network Traffic: The sessions between users/devices and servers are defined by parameters including session duration, protocol type, and packet size.
2. Security Protocols: It is defined by encryption standards (AES, DES) and means of authentication.
3. IDS Sensors: These are the tools monitoring traffic for anomalous behavior, including failed logins, unusual access times.
4. Risk Factors: Vulnerabilities like weak encryption, high-risk protocols (UDP/ICMP), and behavioural anomalies (e.g., brute-force attempts).

A. Assumptions and their justifications

The following assumptions were applied where empirical gaps existed:

1. Session Independence: Sessions were treated as independent trials for Binomial/Poisson modelling. *Reason*: It simplifies analysis despite potential real-world attack dependencies.
2. Fixed Attack Probabilities: Baseline attack rates (e.g., 44.71%) were assumed static. *Reason*: It provides a conservative risk estimate for decision-making.
3. Linearity in Relationships: Variables like session_duration and attack_detected were assumed to be linearly correlated. *Reason*: It facilitates parametric analysis; deviations were addressed via non-parametric tests.

C. Key Challenges

1. **Class Imbalance**: The ratios of attack (44.71%) vs. non-attack (55.29%) risked biased model training.
2. **Censored Data**: 55.29% of sessions lacked attacks, complicating time-to-event analyses.
3. **Non-Numeric Features**: Categorical parameters (e.g., protocol_type) required to be encoded for statistical modelling.

D. Risk Analysis Framework

The dataset enabled the application of:

1. Parametric Models: Poisson (discrete events), Binomial (binary outcomes), Weibull (time-to-attack).
2. Non-Parametric Methods: Kaplan-Meier survival analysis for censored data.
3. Decision-Support Tools: Sensitivity analysis to prioritise cost-effective mitigations (e.g., protocol upgrades).

IV. Risk Assessment

In our dataset, we found two types of parameters, which include numerical and non-numerical data. For the numerical data, distributions such as Poisson, binomial, normal and Weibull were used. Meanwhile, for the non-numeric data, Bayesian analysis was performed.

A. Poisson Distribution

Poisson distribution is used to model the number of events happening in a fixed interval of time or space, given the average number of times the event occurs over that interval. In our dataset, the events are failed login attempts, which makes sense as they are discrete events that can be counted. Also, the failed logins are relatively rare per session, fitting Poisson assumptions and the rate estimation helped to estimate the attack frequency. For instance, What's the chance of ≥3 failed logins? The average rate λ which is the mean number of failed logins per session is found out to be 1.5177728845548915. We, then computed the PMF (probability mass function) and CDF (cumulative distribution function) using the formula: $P(X=k)=\lambda ke-\lambda/k!$ for k from 0 to 8. Then, we calculated the probability of different attack counts which was as follows:

0 Attacks: 21.92%
1 Attack: 33.27%
2 Attacks: 25.25%
3 Attacks: 12.77%
4+ Attacks: Less than 5%

We plotted these probabilities as in fig 5.1 which is the PMF plot displaying the theoretical probability of observing k failed logins. The x-axis shows possible values of k, and the y-axis shows their corresponding probabilities.

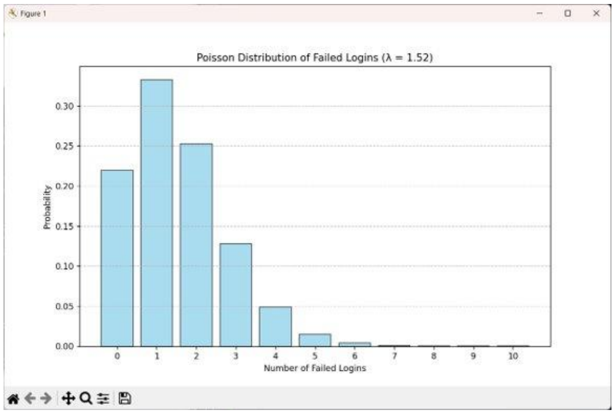


Fig 4.1

Thus, this analysis helped in intrusion detection by identifying anomalous login behavior statistically, whereas the bar chart helped us to visualize how well the Poisson distribution models the frequency of failed login attempts and to detect unusual patterns e.g., too many high k values may indicate brute-force attacks.

B. Binomial Distribution

The Binomial Distribution is used to model the number of successes in a fixed number of independent trials, where each trial has the same probability of success. In our dataset, the attack_detected column is binary (with values 0 or 1), making it suitable for Binomial distribution analysis. We treated each session as a trial, and an attack being detected as a "success." Here, we assumed that each session is independent, there is a constant probability of an attack across all sessions and there can only be Binary outcomes.

Firstly, we calculated the probability of the attack p, where p = Number of Attacks/ Total sessions and probability of no attack q = 1-p. Also, the number of attacks in n sessions follows Binomial(n,p) then the Expected Attacks is given by $\mu=n\cdot p$ and Variance is $\sigma^2=n\cdot p\cdot q$. Calculation results found were as follows:

Total Sessions: 9537
Number of attacks: 4264
Probability of Attack (p) =0.4471
Probability of no attack (q): 0.5529
Expected attacks in next 100 sessions: 44.7
Variance: 24.7
Standard Deviation: 5.0
Probability of >10 attacks in 50 sessions: 0.9998

Here, Higher p means greater attack risk. Example we computed p = 0.4471, which implies there's a 44.71% chance of an attack per session and expected attacks in next 100 sessions: 44.7

Key Findings from the Dataset using Binomial Distribution was:

- `protocol_type`: UDP/ICMP have higher attack rates.
- `failed_logins`: Sessions with more failed login attempts may indicate brute-force attacks.
- `encryption_used`: Weak encryption (e.g., DES) correlates with higher attack success.
- `ip_reputation_score`: Lower scores indicate malicious IPs.

We obtained 3 main plots: a count plot of attack occurrences, a binomial distribution PMF, and a correlation heatmap.

1. Attack Distribution (Count Plot) and Binomial Distribution (PMF Plot)

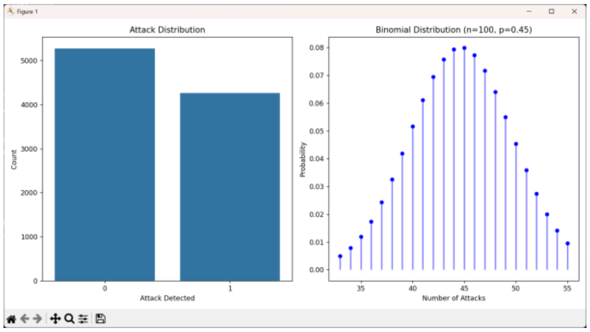


Fig. 4.2

The bar chart (in Fig. 5.2) compares the number of sessions with attacks (attack_detected=1) vs. no attacks (attack_detected=0). A probability mass function (PMF) plot of the expected number of attacks in future_sessions (default: 100 sessions).

Blue dots represent probabilities for different attack counts.

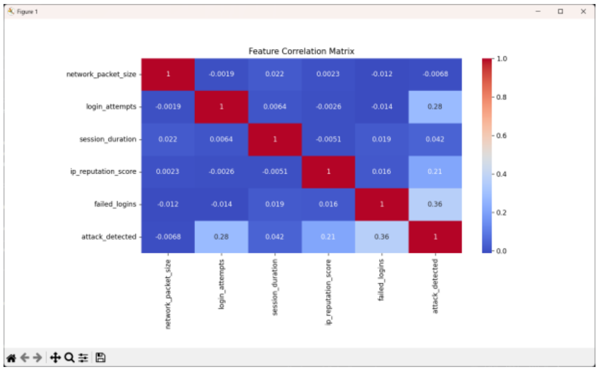
Key Insights: The bar chart reveals the baseline attack rate in the dataset (e.g., 30% attacks vs. 70% normal traffic) and helps assess the class imbalance. The PMF plot predicts the likelihood of specific attack counts (e.g., "There's a 20% chance of 25 attacks in 100 sessions") and shows the spread of risk (mean ± standard deviation) around expected attacks.

2. Feature Correlation Heatmap

This color-coded matrix (Fig. 5.3) shows correlations between numeric features (e.g., failed_logins, session_duration) and attack_detected.

Key Insights: It identifies risk factors (e.g., high failed_logins may correlate with attacks). It also highlights protective factors (e.g., high ip_reputation_score may reduce attack likelihood).

Here, red = positive correlation, Blue = negative correlation.



	network_packet_size	login_attempts	session_duration	ip_reputation_score	failed_logins	attack_detected
network_packet_size	1	-0.0019	0.022	0.0023	-0.012	-0.0068
login_attempts	-0.0019	1	0.0064	-0.0026	-0.014	0.28
session_duration	0.022	0.0064	1	-0.0051	0.019	0.042
ip_reputation_score	0.0023	-0.0026	-0.0051	1	0.016	0.21
failed_logins	-0.012	-0.014	0.019	0.016	1	0.36
attack_detected	-0.0068	0.28	0.042	0.21	0.36	1

Fig. 4.3

C. Normal Distribution

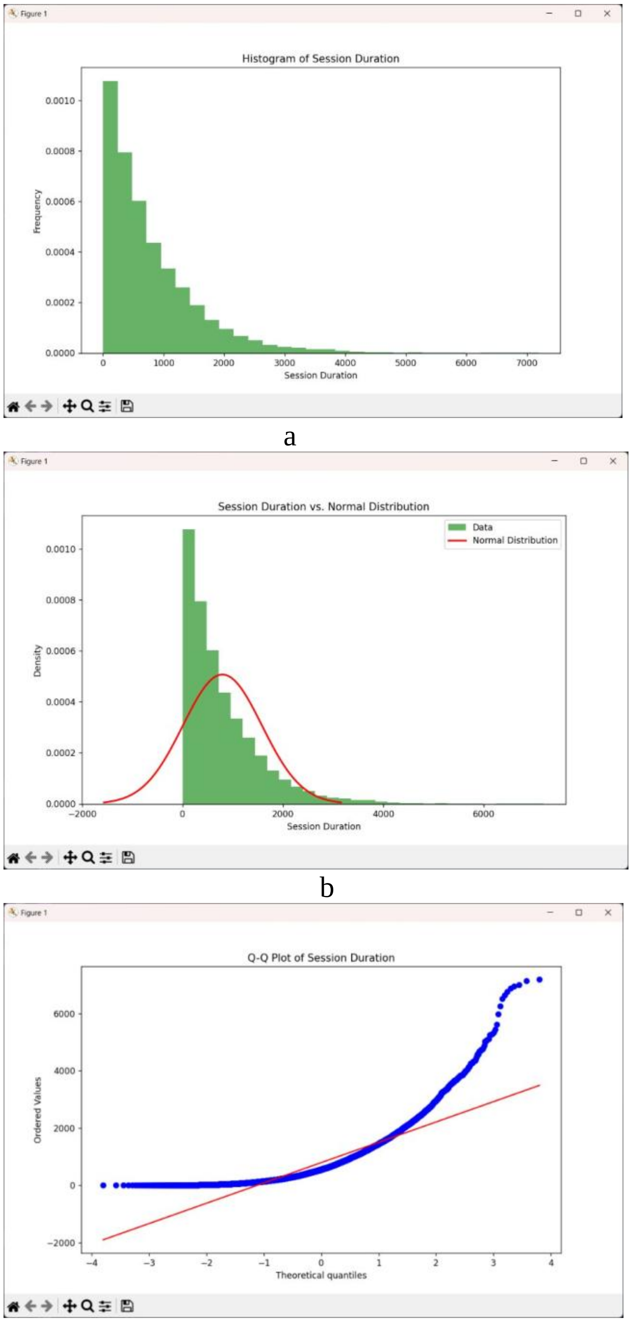
It is a continuous probability distribution that is symmetric around the mean and shows that data near the mean are more frequent in occurrence than data far from the mean. We selected the variable: session_duration which is continuous for this analysis and calculated mean and standard deviation. Then, we plotted the histogram or Q-Q plot and performed the normality test (Shapiro-Wilk test). If it is normal, use Z-scores for probabilities and if not, mention the deviation. In cybersecurity, knowing the distribution helps set thresholds for anomalies (e.g., flagging sessions with durations in the top 5% as suspicious).

Here, we assume that the observations/sessions are independent of each other and extreme values (e.g., abnormally long sessions) do not skew the distributions. Also, the dataset size (1,400 entries) is large enough for the Central Limit Theorem to apply and relationships between variables (if analyzed) are linear.

Normal Distribution Results:
Mean: 792.75
Standard Deviation: 786.56
Shapiro-Wilk p-value: 0.0000 (Data is NOT normally distributed). A Shapiro-Wilk p-value > 0.05 suggests normality.

We created a histogram with a KDE line, a Q-Q plot, and overlaying a normal distribution curve on the histogram. Each of these graphs serves a specific purpose in assessing normality. The histogram (in Fig. 5.4(a)) shows the distribution of the data, such as session duration. If it is bell-shaped, that is a sign of normality.

The KDE line helps visualize the density. The Q-Q plot(c) compares the data quantiles to a theoretical normal distribution and if these points lie roughly on the red line, the data is normal. Deviations, especially at the ends, indicate skewness or outliers. The overlay of the normal curve on the histogram (b) allows a direct comparison.



a

b

c

Fig. 4.4

Key Takeaways for our Dataset

- session_duration is likely right-skewed (short sessions dominate, few extremes).
- network_packet_size is possibly bimodal (mixed traffic types).
- ip_reputation_score may cluster near 0 (low reputation) with outliers near 1.

D. Weibull Distribution

Weibull Analysis is a statistical method used for life data analysis and helps in understanding the failure rates over time. It is used to model time-to-event (attack detection time) and is defined by two parameters: β (shape) and α (scale). The shape parameter β tells us about the failure rate over time and the scale parameter is the characteristic life, where 63.2% of units are expected to have failed. If β < 1, the failure rate decreases over time; β=1 means constant rate (exponential distribution); β>1 means increasing failure rate.

- From the Weibull analysis of the dataset and using columns session_duration (time until an attack is detected, or session ends without an attack), attack_detected (Binary indicator, where 1 = attack, 0 = no attack), we obtained the values of β (shape) = 1.02, α (scale) = 1756.62. Since β ≈ 1, the risk (hazard rate) remains almost constant over time. So, after about 1757 seconds (≈29 minutes), most attacks have already taken place. Also, it is Right-Censored Data (55.29%) which implies 55.29% of the sessions ended without an attack (censored data).
- Possible Security Strategies – Expert Opinion
  1. General Continuous Monitoring - Risk is constant, so an attack can occur at any time.
  2. Authentication & Initial Security Measures: Focus on strong authentication at the start of a session to prevent unauthorized access early.
  3. Session Timeout Consideration: Unlike a scenario where attacks increase over time (β > 1), setting a hard session limit (like 600 seconds) may not significantly reduce risk in this case.

E. Survival Analysis - The Kaplan-Meier estimator

Survival Analysis is used to analyse time-to-event data, like how long until an event happens. The Kaplan-Meier estimator is a common method for this, which estimates the survival function based on observed times.

The key elements for survival analysis are:

1. Time-to-event: This is the duration until the event occurs. In this dataset, session_duration seems like the time column. It represents how long the session lasted.

2. Event indicator: This indicates whether the event (in this case, an attack being detected) occurred. The attack_detected column is binary (0 or 1), which fits this requirement. Other columns like network_packet_size, protocol_type, etc., might be covariates, but for the basic Kaplan-Meier analysis, we just need the time and event indicator.

Then, the data was sorted based on session duration in ascending order, and events (rows where attack_detected = 1) and censored data (rows where attack_detected = 0) were separated before computing the Kaplan-Meier Survival Curve. The figure (Fig. 5.5) below displays the graph obtained from the dataset assessment.

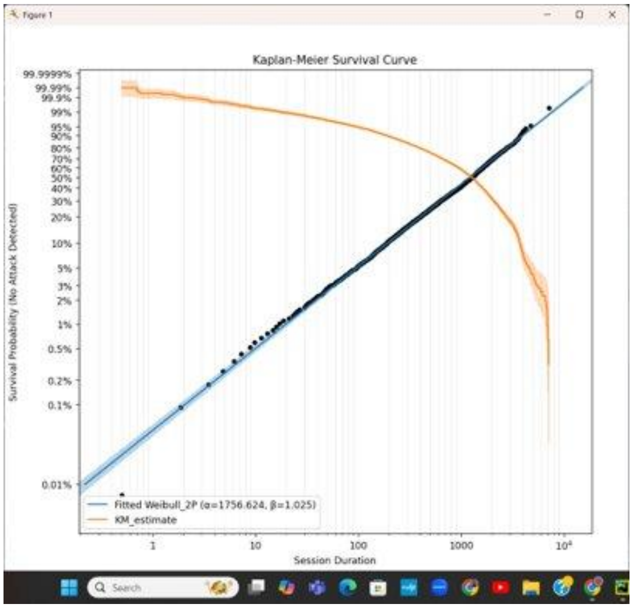


Fig. 4.5

The X-axis (SessionDuration) represents the elapsed time (in seconds, based on the dataset) since the session began.

And the Y-axis (Survival Probability) shows the probability that a session remains attack-free up to a given time. It starts at 1 (100% probability of no attack at time 0) and decreases over time as attacks occur.

The initial Steep drop in the curve suggests that high-risk periods occur shortly after sessions begin. Example: If many attacks occur within the first 100 seconds, the survival probability drops rapidly here.

The final survival probability (e.g., S(t)=0.2*S*(*t*)=0.2 at t=5000*t*=5000) shows the **overall resilience** of sessions.

Key takeaways

- The Kaplan-Meier analysis shows how long sessions remain safe from attacks.
- If survival probability drops quickly, it means sessions are at high risk of attack early. If the probability declines gradually, risk is more evenly distributed over time.
- Based on this, you can make data-driven security decisions like setting session limits, improving early security measures, or focusing monitoring efforts at specific times.

F. Bayesian Analysis

To analyze and quantify the risk level based on observed risk patterns, including encryption algorithms, network protocol types, browser use, and unusual access, we used the Bayes algorithm.

$$P(h/e) = \frac{p(e/h)\ p(h)}{P(e)}$$

Where,
P(e) - Given that different parameters have taken place
P(h) - hypothesis that attack has occurred given prior event

Total events	9537
Feature	**Count**
Attack	4264
attack when TCP used	2963
UDP used and attack	1092
AES used and attack	2055
DES used and attack	1299
Chrome used and attack	2202
Firefox used and attack	849
unusual time of access and attack	654
non-attack using TCP	3661
non-attack using UDP	1314
AES and non-attack	2651
DES and non- attack	1566
Chrome non-attack	2935
Firefox non-attack	1095
unusual time of access non-attack	776

Table 4.1

Feature (X)	P (Attack | X)	Effect on Attack Probability
TCP	44.70%	No effect
UDP	45.40%	Slight increase (+0.7%)
AES	43.70%	Slight decrease (-1.0%)
DES	45.50%	Slight increase (+0.8%)
Chrome	42.80%	Decrease (-1.9%)
Firefox	43.60%	Decrease (-1.1%)
Unusual time	45.60%	Slight increase (+0.9%)

Table 4.2

From table 1.2 we see those parameters like UDP protocol usage, DES encryption algorithm usage and unusual access of servers shows higher risk patterns when compared to their alternatives of using AES and TCP. Although we also notice that the type of browser including chrome and Firefox do not contribute to high risk but rather lower the risk of an attack.

V. RISK ANALYSIS

To perform risk analysis from the identified risk patterns, we used Fault tree analysis, Event tree analysis, sensitivity analysis, Bayesian Network analysis, and Payoff tables.

A. Fault tree Analysis

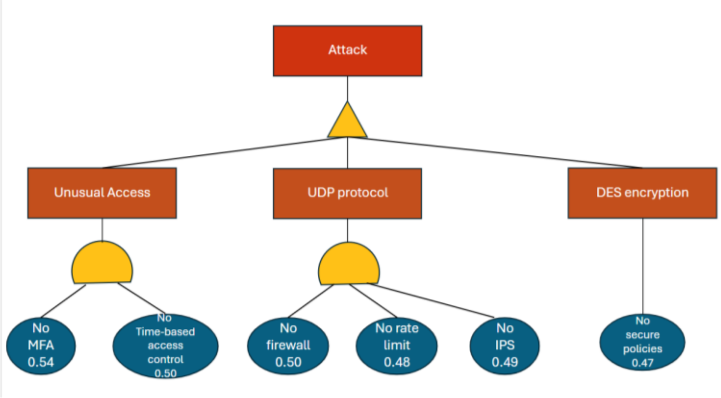


Fig 5.1

From the fault tree, we see that a high attack probability is possible even if one of the main components, protocol type, encryption algorithm, or no protection against unusual access fails. Considering the individual components, Failure of the system can occur if there is no multi-factor authentication (MFA) and no time-restricted access during unusual hours. Similarly, for the protocol type, even if the organization chooses to use UDP instead of TCP, the attack probability increases when there is no firewall protection, IPS, or rate limit. Using DES encryption instead of AES, along with no security policies, could directly affect the system failure probability.

B. Event tree analysis

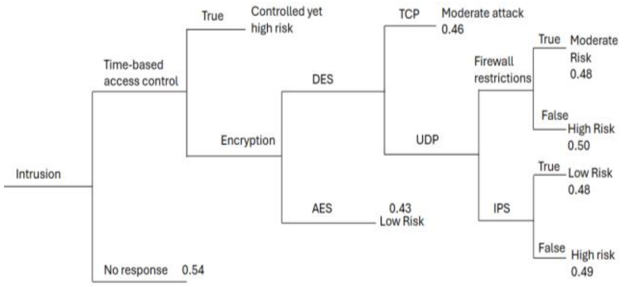


Fig 5.2

From the event tree, we notice that having no response to an intrusion has the highest probability of an attack. However, use of a time-based control system prevents unusual access to the network but does not protect us from other attacks, hence it controls the attack surface but still poses a high risk.

Using AES for encryption shows the lowest probability of attack, ensuring a safe network. Whereas, using DES along with UDP with the least firewall restriction poses the highest risk factor when compared to using TCP or using UDP with an Intrusion Prevention System (IPS).

When an intrusion occurs, the best possible response to prevent an attack would be to use proper encryption algorithms and protocols along with restricted access. This strategy would ensure the lowest risk probability for an attack to succeed.

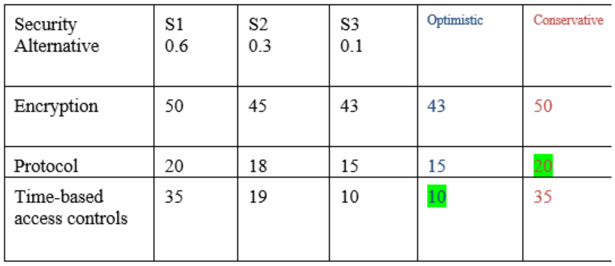
C. Sensitivity Analysis

The analysis was performed with the following assumptions.

1. An organization holding an IDS wants to enhance the security of its network.
2. The organization has alternatives of enhancing encryption algorithms from using DES to AES, enhancing the protocol to use TCP compared to UDP, and implementing MFA and time-restricted access.
3. Based on prior risk probabilities for the organization to face high risk(S1=0.60), medium risk(S2=0.30), and low risk(S3=0.10) in the next decade is assumed.
4. The cost of implementing these security measures under different states of nature is assumed.

Starting with the optimistic approach to minimize (minimin) the cost of expenditure, based on calculations performed, the organization should consider implementing time-based risk control measures.

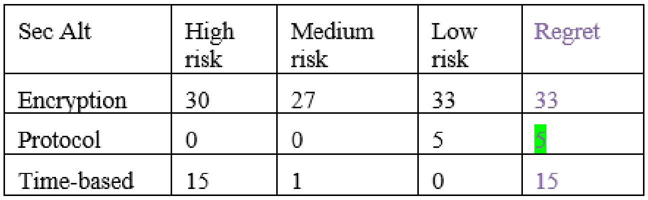
Based on a conservative approach to saving costs on security, the organization would have to consider implementing protocol enhancements. This could be ideal if the organization has medium to low-risk predictions.



Security Alternative	S1 0.6	S2 0.3	S3 0.1	Optimistic	Conservative
Encryption	50	45	43	43	50
Protocol	20	18	15	15	20
Time-based access controls	35	19	10	10	35

Table 5.1

Considering the regret table to decide on the security measures by minimizing the maximum regret for the organization, the best decision would be to enhance protocols.



Sec Alt	High risk	Medium risk	Low risk	Regret
Encryption	30	27	33	33
Protocol	0	0	5	5
Time-based	15	1	0	15

Table 5.2

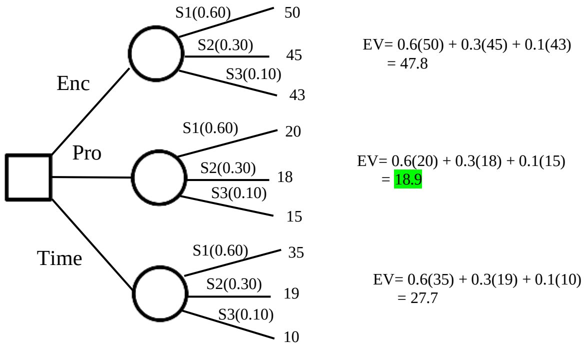


Fig 5.3

Based on the decision tree, we have calculated the Expected Value (EV) for each security alternative. The EV of the decision tree is to enhance protocols given the low cost of implementation.

To perform sensitivity analysis, assuming the probability changes,

P(s1) =p

P(s2) =0.3 x (1-p)/0.4

P(s3) =0.1 x (1-p)/0.4

After plotting the probabilities and expected costs in the X and Y axis respectively, we can conclude that, for high-risk scenarios, investing in proper encryption algorithms would benefit from causing huge loss due to attacks on the organization. Although investing in efficient protocol types like TCP is not only cost efficient but would be the optimistic decision given the changes in the states of nature whereas encryption optimization would not be cost efficient if risk level for the organization decreases.

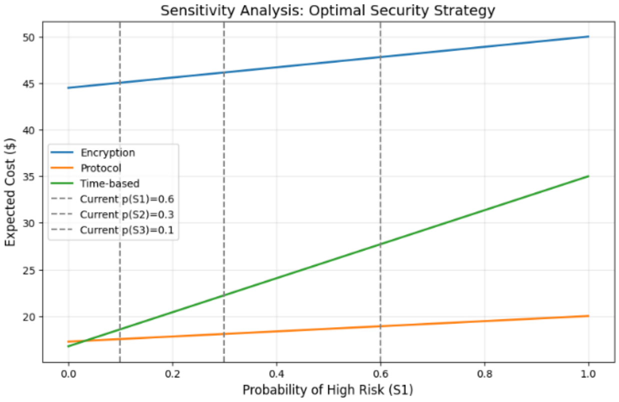


Fig 5.4

D. Bayesian Network

In this analysis of Bayesian network, we analyze the conditional dependence and how they influence or increase the risk of an attack.

For the analysis in Fig 4.4 we have used the parameters,

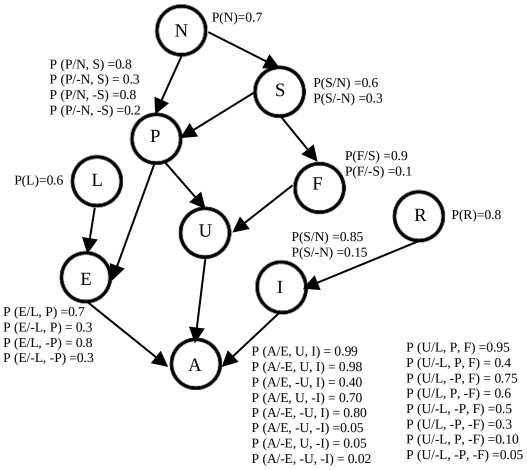


Fig 5.5

N- Network packet size, P- Protocol used, S- Session duration, L- Login attempts, F-failed logins, U- Unusual access, R- Request anomalies, I- IPS alerts, E- Encryption used, A- Attack detected

From the network we see,

$$P(N,S,P,L,F,R,U,E,I,A)=P(N)P(L)P(R)P(S/N)P(P/N,S)$$

$$P(F/S)P(U/L,P,F)P(E/L,P)P(I/R)P(A/E,U,I)$$

$$=(0.7)(0.6)(0.8)(0.6)(0.8)$$

$$(0.9)(0.95)(0.7)(0.85)(0.99)$$

$$=0.0812$$

Given the low value of joint probability calculated, we can say that this combination of attacks is rare yet possible.

From the Bayesian network some key findings include,

i. The Encryption algorithm, unusual access and IPS alerts play a major role in the attack surface where they spike the attack probability to 99% if not secure.

ii. The probability of an unusual access to the organization's network depends on various factors including failed logins, the protocol type used and the number of login attempts.

iii. Although IPS alerts can directly cause an attack to be detected, it is also in direct proportion to the number of Requests received from the malicious node.

VI. Discussion

The comprehensive risk analysis performed gave us a layered approach to understand the intrusion risk patterns and mitigations. Fault tree and event tree analysis gave an attack pathway, mapping and highlighting the high-risk combinations to be avoided. Using Decision tree analysis, we were able to figure out cost-effective approaches depending on the risk probabilities specific to organizations need.

Sensitivity Analysis helped us determine the effectiveness of certain security alternatives and the dominant risk outcome of other parameters based on constant changes to the states of nature. Bayesian Analysis helped carve out a clear picture of inter dependent features and risk factors that contribute to an attack being intensely categorized as high-risk to low-risk combination of attacks.

Limitations: The data and assumptions used for this study is taken from one IDS system and can be biased. The results can vary depending on the organization and type of IDS used for detection.

Future Directions: A good approach would be to implement risk analysis using dynamic data streams to update probabilities real-time resulting in accurate predictions. Another suggestion for future research would be to use automated tools for threat modelling that could generate CPTs from threat intelligence feeds.

VII. Conclusion

This study conducted a comprehensive risk analysis leveraging data from Intrusion Detection Systems (IDS) to identify critical cybersecurity vulnerabilities and inform actionable mitigation strategies. By integrating statistical methodologies—including Poisson, Binomial, Normal, and Weibull distributions—the research modelled discrete and continuous risk events such as failed logins, attack probabilities, session durations, and time-to-attack patterns. Bayesian analysis further quantified conditional dependencies between risk factors, while decision-support tools like Fault Tree Analysis (FTA) and Sensitivity Analysis evaluated the efficacy of security measures.

Key findings underscored that protocol types (e.g., UDP), weak encryption standards (e.g., DES), and unusual access times significantly heightened attack risks. For instance, sessions using UDP exhibited a 45.4% attack probability compared to TCP's 44.7%, and DES encryption correlated with a 45.5% attack likelihood versus AES's 43.7%. Survival analysis via the Kaplan-Meier estimator revealed that over 50% of attacks occurred within the first 29 minutes of a session, emphasizing the need for robust initial security measures.

Decision frameworks highlighted protocol enhancements and encryption upgrades as cost-effective strategies to mitigate risks.

By bridging statistical rigor with practical cybersecurity decision-making, this work provides organizations with a replicable methodology to prioritize defences, optimize resource allocation, and transition from reactive to proactive risk management. Future research should validate these approaches across diverse network environments and incorporate dynamic threat modelling to address evolving cyberattack tactics.

References

[1] D. Kumar, "Cybersecurity Intrusion Detection Dataset," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/dnkumars/cybersecurity-intrusion-detection-dataset.

[2] I. H. Sarker et al., "Cyber Risk Learning from Data: A Review," *IEEE Access*, vol. 8, pp. 191474–191491, 2020, doi: 10.1109/ACCESS.2020.3032250.

[3] N. Fenton and M. Neil, *Risk Assessment and Decision Analysis with Bayesian Networks*. CRC Press, 2012.