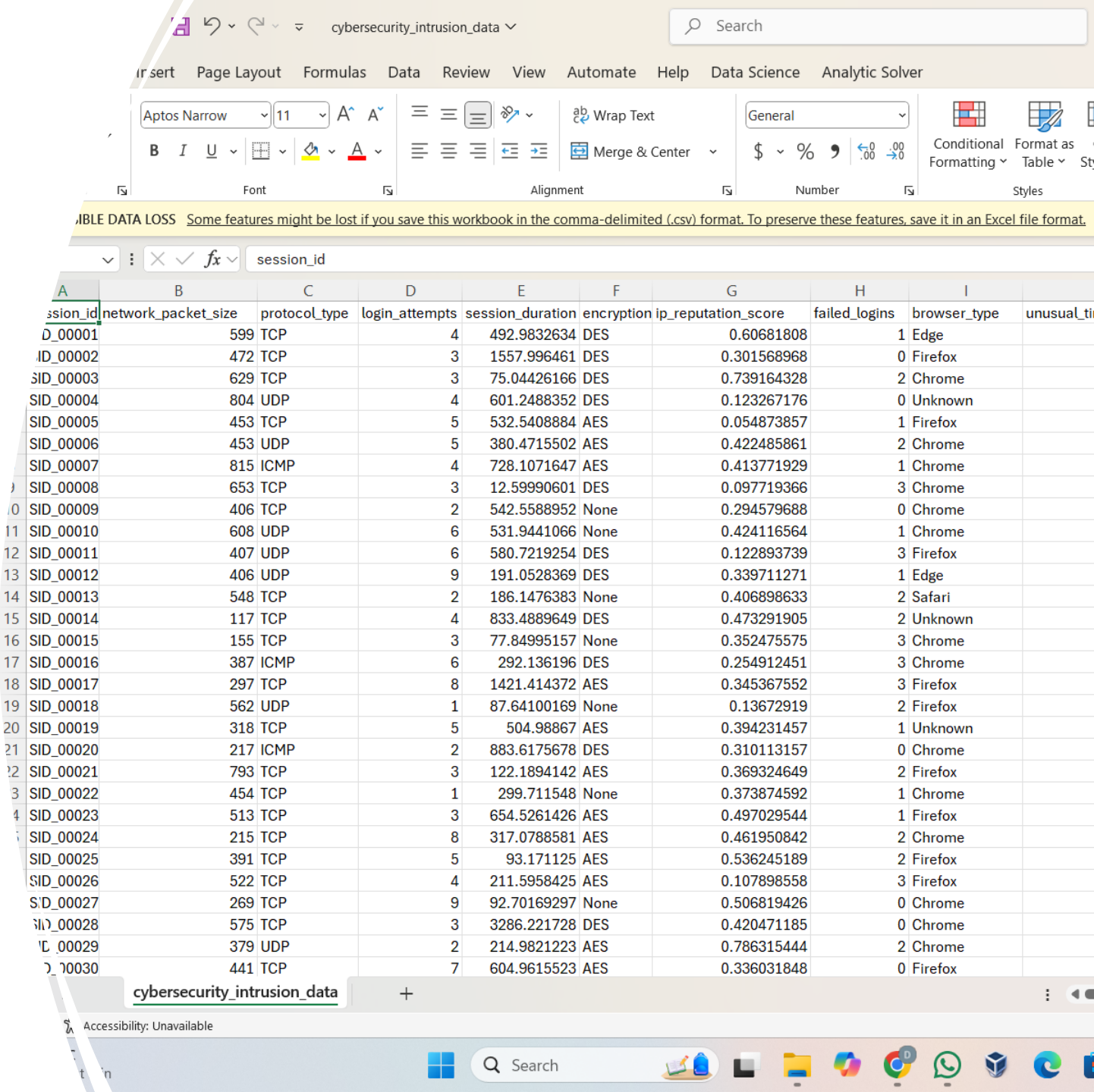
Risk Analysis On Cyber Attacks: Evaluating Success Probability and Detection Accuracy

Dataset

- We chose to collect data from an IDS
- Track network packet features, analyze risk in each of them

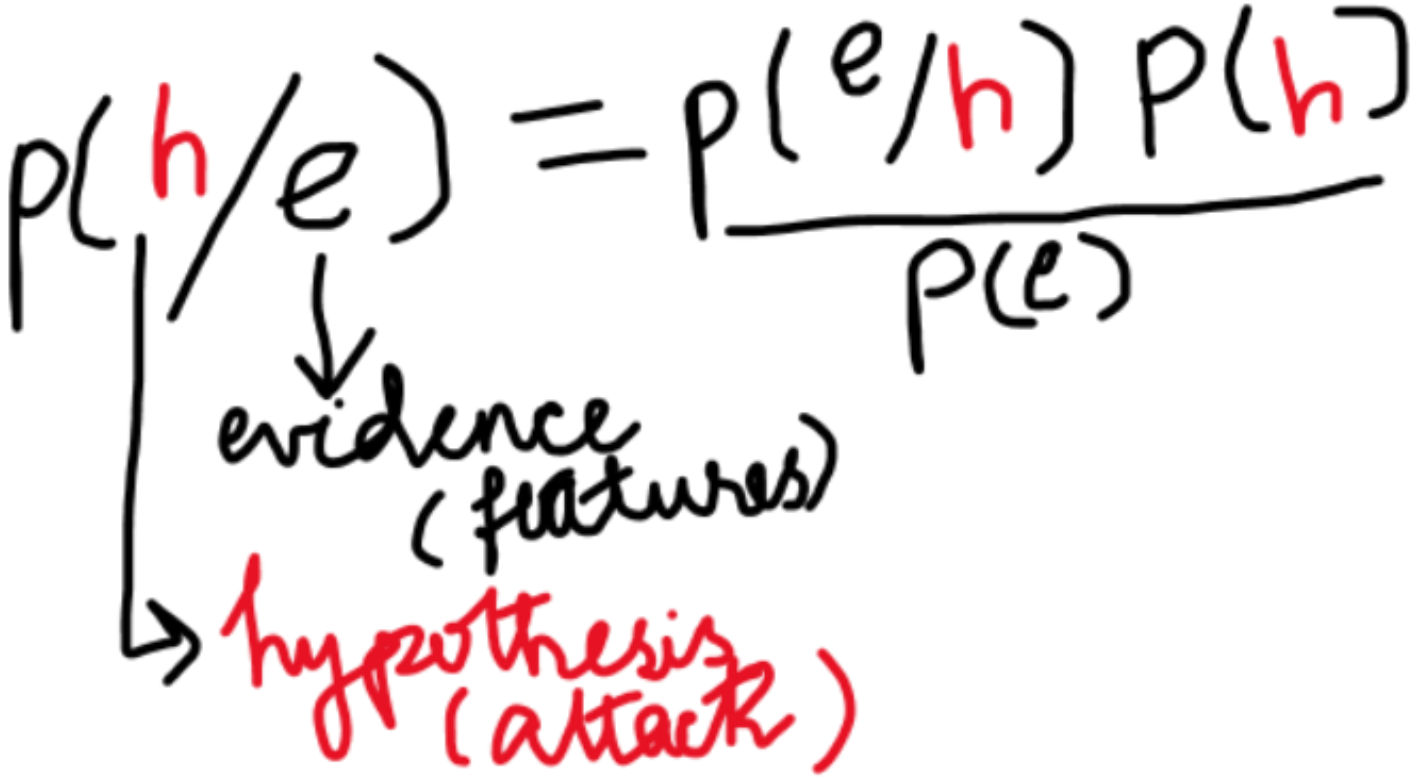


cybersecurity_intrusion_data

session_id	network_packet_size	protocol_type	login_attempts	session_duration	encryption	ip_reputation_score	failed_logins	browser_type	unusual_ti
D_00001	599	TCP	4	492.9832634	DES	0.60681808	1	Edge	
ID_00002	472	TCP	3	1557.996461	DES	0.301568968	0	Firefox	
SID_00003	629	TCP	3	75.04426166	DES	0.739164328	2	Chrome	
SID_00004	804	UDP	4	601.2488352	DES	0.123267176	0	Unknown	
SID_00005	453	TCP	5	532.5408884	AES	0.054873857	1	Firefox	
SID_00006	453	UDP	5	380.4715502	AES	0.422485861	2	Chrome	
SID_00007	815	ICMP	4	728.1071647	AES	0.413771929	1	Chrome	
SID_00008	653	TCP	3	12.59990601	DES	0.097719366	3	Chrome	
SID_00009	406	TCP	2	542.5588952	None	0.294579688	0	Chrome	
SID_00010	608	UDP	6	531.9441066	None	0.424116564	1	Chrome	
SID_00011	407	UDP	6	580.7219254	DES	0.122893739	3	Firefox	
SID_00012	406	UDP	9	191.0528369	DES	0.339711271	1	Edge	
SID_00013	548	TCP	2	186.1476383	None	0.406898633	2	Safari	
SID_00014	117	TCP	4	833.4889649	DES	0.473291905	2	Unknown	
SID_00015	155	TCP	3	77.84995157	None	0.352475575	3	Chrome	
SID_00016	387	ICMP	6	292.136196	DES	0.254912451	3	Chrome	
SID_00017	297	TCP	8	1421.414372	AES	0.345367552	3	Firefox	
SID_00018	562	UDP	1	87.64100169	None	0.13672919	2	Firefox	
SID_00019	318	TCP	5	504.98867	AES	0.394231457	1	Unknown	
SID_00020	217	ICMP	2	883.6175678	DES	0.310113157	0	Chrome	
SID_00021	793	TCP	3	122.1894142	AES	0.369324649	2	Firefox	
SID_00022	454	TCP	1	299.711548	None	0.373874592	1	Chrome	
SID_00023	513	TCP	3	654.5261426	AES	0.497029544	1	Firefox	
SID_00024	215	TCP	8	317.0788581	AES	0.461950842	2	Chrome	
SID_00025	391	TCP	5	93.171125	AES	0.536245189	2	Firefox	
SID_00026	522	TCP	4	211.5958425	AES	0.107898558	3	Firefox	
SID_00027	269	TCP	9	92.70169297	None	0.506819426	0	Chrome	
SID_00028	575	TCP	3	3286.221728	DES	0.420471185	0	Chrome	
SID_00029	379	UDP	2	214.9821223	AES	0.786315444	2	Chrome	
SID_00030	441	TCP	7	604.9615523	AES	0.336031848	0	Firefox	

Identify and Access Risks

Bayesian Analysis



Bayesian Analysis

$$P(\text{attack} / \text{unusual time}) = \frac{P(U/A)\,P(A)}{P(U)}$$

$$P(U) = \frac{654 + 776}{9537} = \frac{1430}{9537} = 0.15$$

$$= \frac{0.153\ (0.447)}{0.15}$$

$$= 0.456\ (45.60\%)$$

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

Bayesian Analysis

P(Attack)= 44.70%

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%)

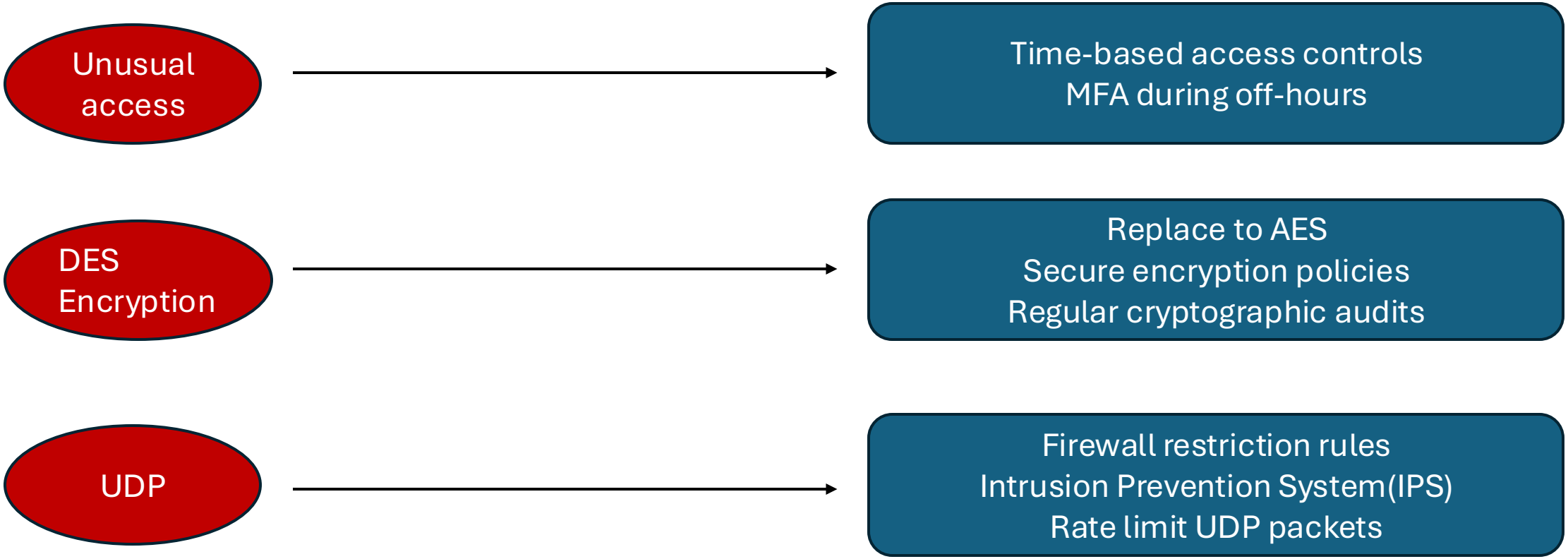
Risk Analysis

P (Attack | X)	Risk Level	Interpretation
>50%	**High Risk**	Attack is more likely than not
40-50%	**Moderate Risk**	Attack is somewhat likely
<40%	**Low Risk**	Attack is less likely

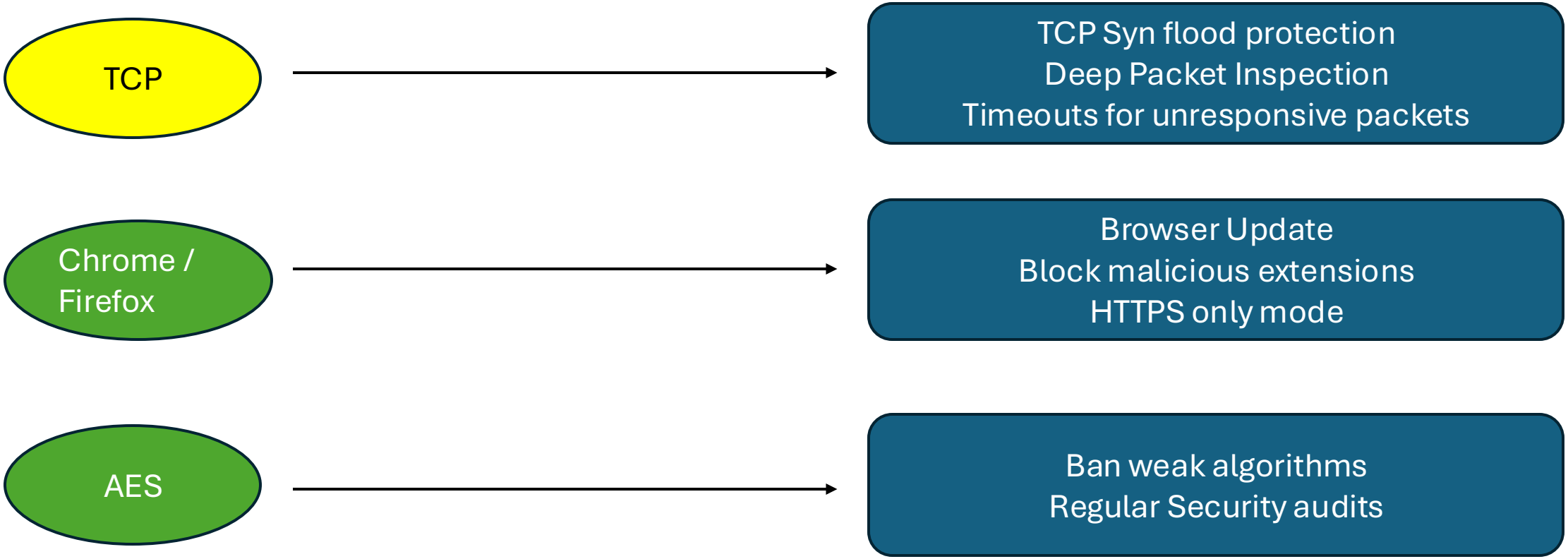
Risk Analysis

Feature	Risk percentage	Risk level	Mitigation Priority
Unusual access time	45.60%	High	Immediate action needed
DES Encryption	45.50%	High	Immediate action needed
UDP usage	45.40%	High	Immediate action needed
TCP Usage	44.7%	Neutral	Recommended but not immediate
Chrome & Firefox	42.8% & 43.6%	Low	Low priority
AES Encryption	43.7%	Low	Safe state

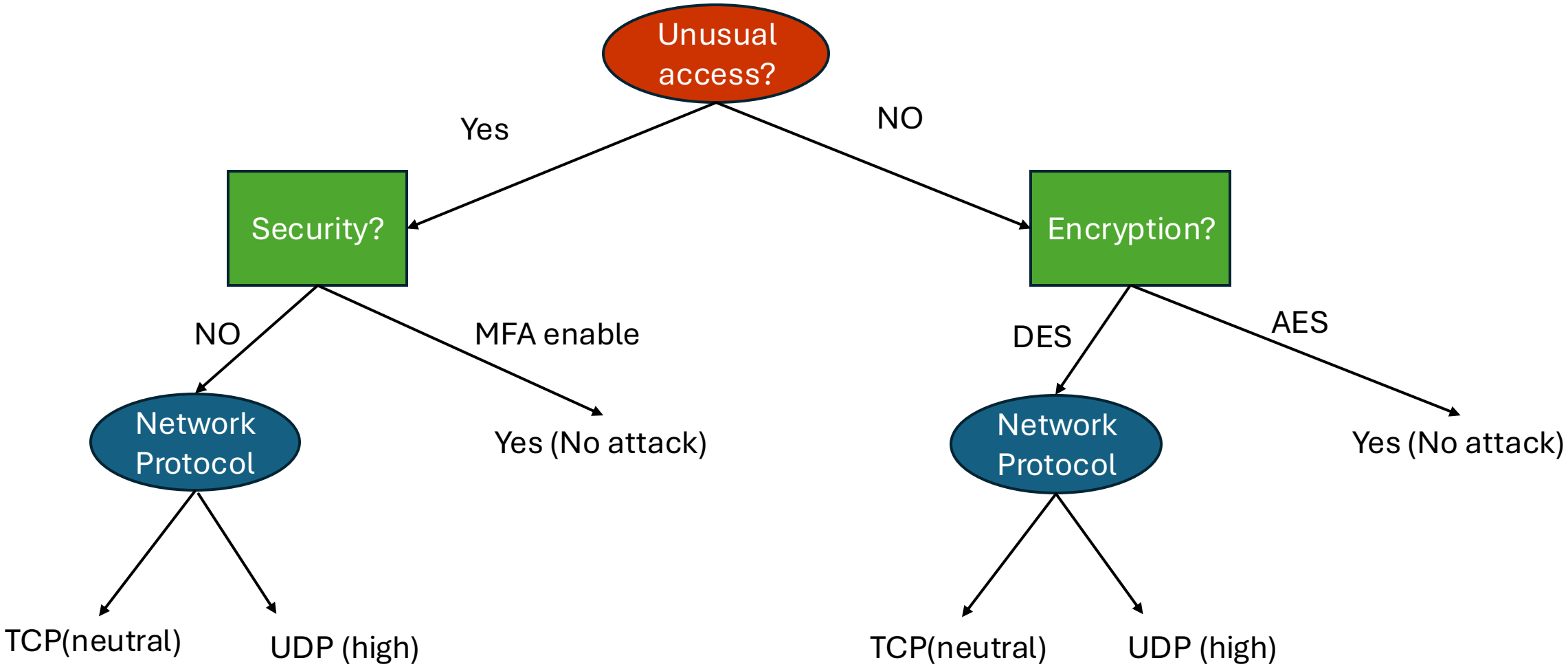
Risk Response and Countermeasures



Risk Response and Countermeasures



Decision Tree



SECURITY MECHANISMS

Implement time based access control

High risk environment

Low risk env

Medium risk env

Enhance Encryption algorithm

High risk environment

Low risk env

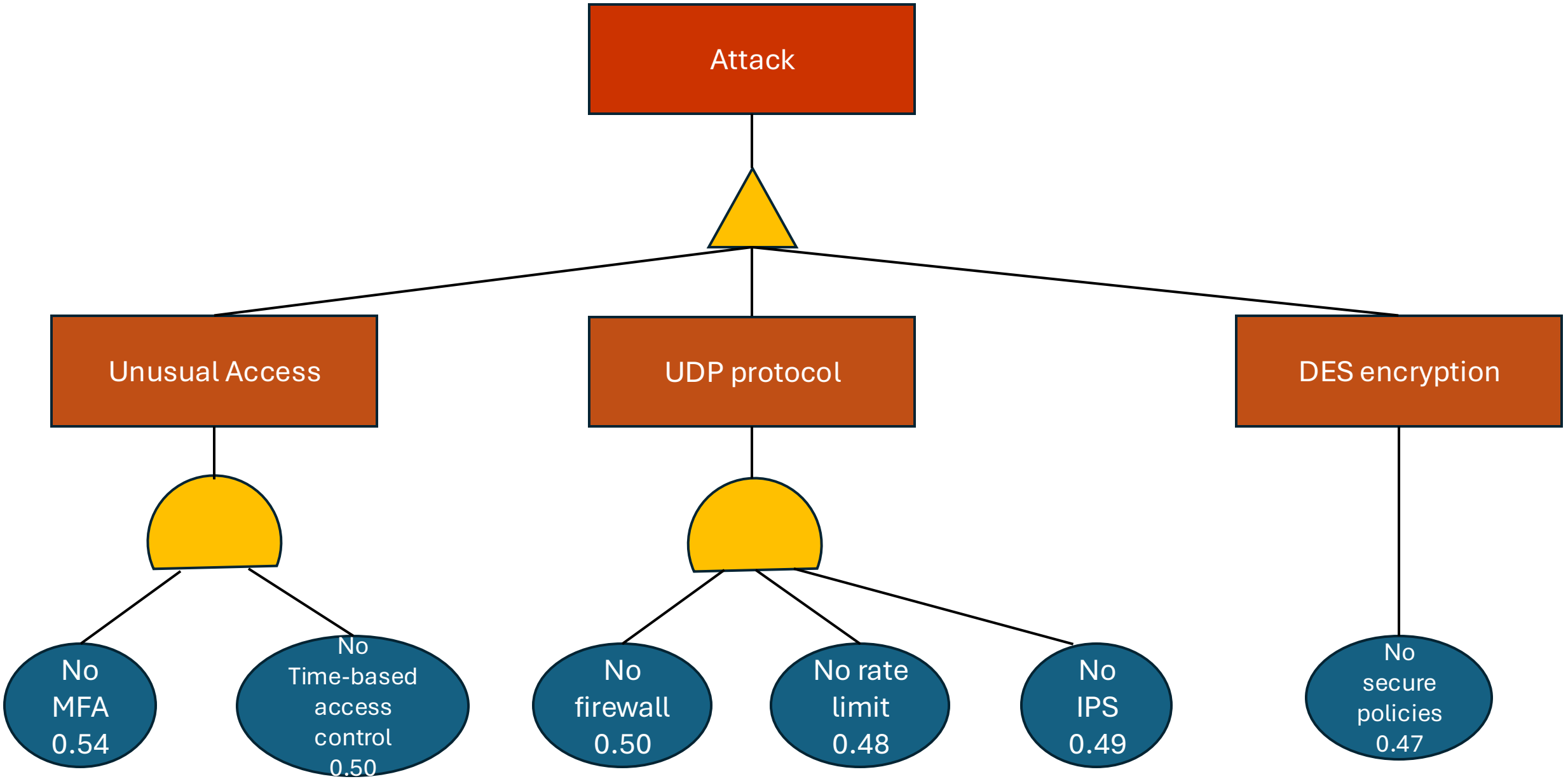
Medium risk env

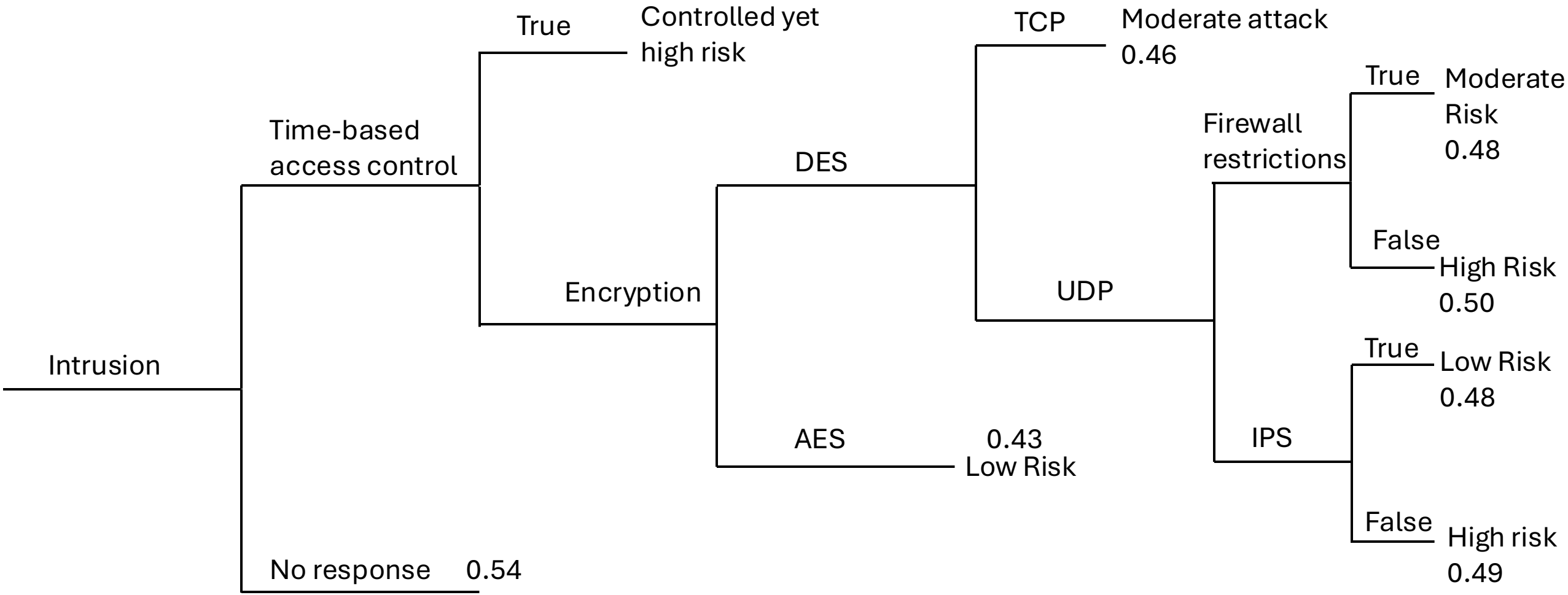
Enhance Protocol type

High risk environment

Low risk env

Medium risk env


Attack
Unusual Access
UDP protocol
DES encryption
No MFA 0.54
No Time-based access control 0.50
No firewall 0.50
No rate limit 0.48
No IPS 0.49
No secure policies 0.47


Intrusion
Time-based access control
No response
0.54
True
Controlled yet high risk
Encryption
DES
AES
0.43
Low Risk
TCP
Moderate attack
0.46
UDP
Firewall restrictions
True
Moderate Risk
0.48
False
High Risk
0.50
IPS
True
Low Risk
0.48
False
High risk
0.49

Expert Opinion

- Risk Analysis
  - Unusual time access + no MFA causes high risk
  - DES + UDP causes high impact Risk
  - Having some response will likely reduce attack surface
- Control Measures
  - Mandatory MFA on all logins
  - DES is outdated and vulnerable, Update to AES
  - Intrusion Prevention System should be implemented to reduce attack %
  - Regular Security audits to revise security controls.

Poisson Distribution

- Poisson distribution is used to model the probability of a given number of events occurring in a fixed interval.
- To find probabilities for different intrusion counts and gain insights into the dataset's behavior.
- λ (Average Rate) = 1.5177 where

λ=Total number of failed logins/Total number of sessions

main.py × execute.py

```
11  import matplotlib.pyplot as plt
12
13  # Load the dataset
14  df = pd.read_csv(r'C:\Users\Manraj Kaur\Desktop\6320 project\dataset... intrusion detection data.csv')
15
16  # Extract 'failed_logins' and calculate lambda
17  failed_logins = df['failed_logins'].tolist()
18  lambda_ = sum(failed_logins) / len(failed_logins)
19  print(f"Average failed logins per session (λ): {lambda_:.2f}")
20
21  # Generate PMF values for k = 0 to 10 (adjust range as needed)
22  k_values = range(0, 11)
23  pmf = [poisson.pmf(k, lambda_) for k in k_values]
24
25  # Plot the Poisson distribution
26  plt.figure(figsize=(10, 6))
27  plt.bar(k_values, pmf, color='skyblue', edgecolor='black', alpha=0.7)
28  plt.title(f'Poisson Distribution of Failed Logins (λ = {lambda_:.2f})')
29  plt.xlabel('Number of Failed Logins')
30  plt.ylabel('Probability')
31  plt.xticks(k_values)
32  plt.grid(axis='y', linestyle='--', alpha=0.7)
33  plt.show()
34
35  print("End of poisson distribution"
```

Poisson Distribution Results

Probability of different attack counts:
0 Attacks: 21.92%
1 Attack: 33.27%
2 Attacks: 25.25%
3 Attacks: 12.77%
4+ Attacks: Less than 5%

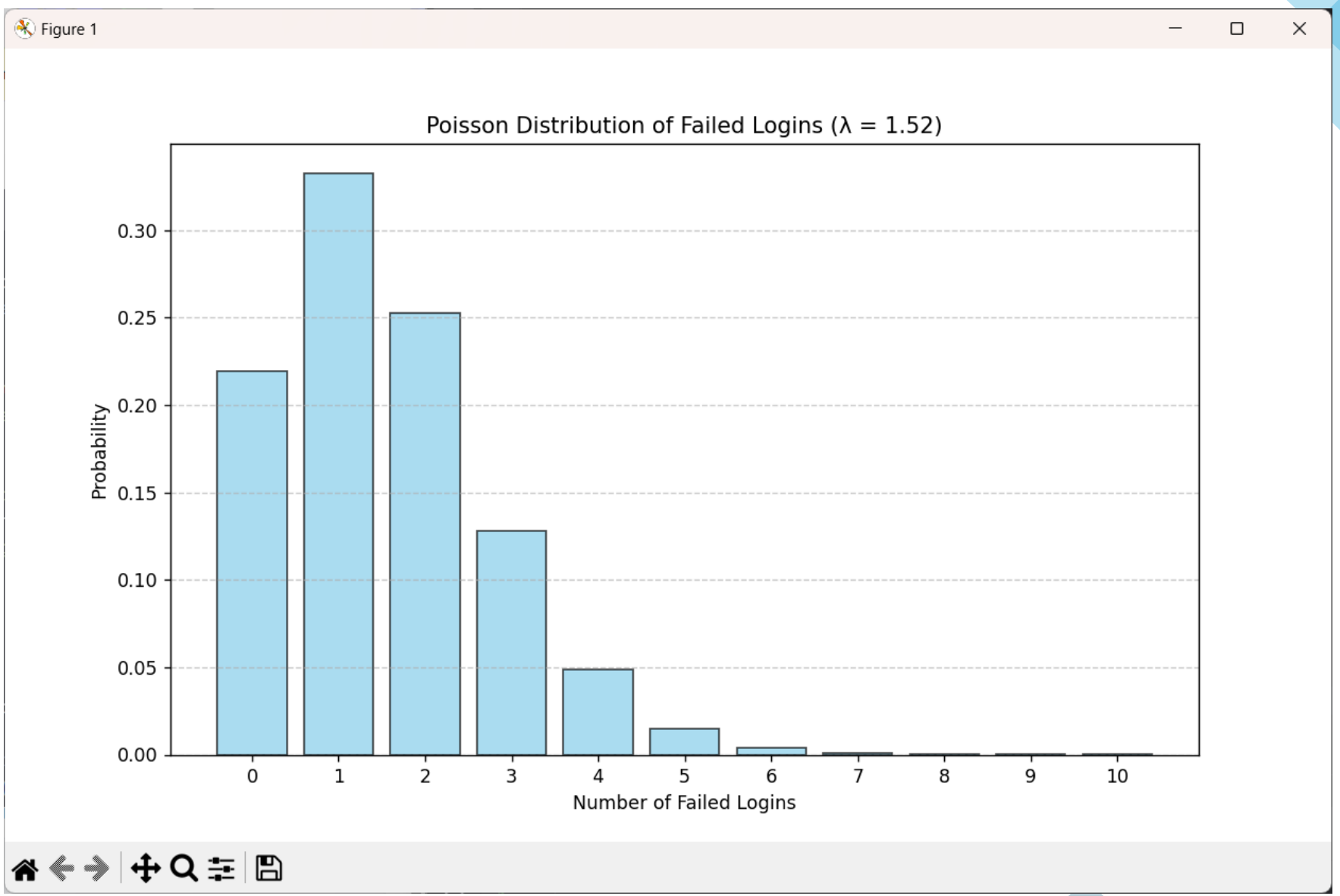


Figure 1
Poisson Distribution of Failed Logins (λ = 1.52)
Probability
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
1
2
3
4
5
6
7
8
9
10
Number of Failed Logins

Binomial Distribution Setup

- Assumptions:
- - Independent trials (each session is independent)
- - Constant probability of attack
- - Binary outcome (Attack = 1, No Attack = 0)

Binomial Distribution Calculation results

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

Risk Analysis

Higher p means greater attack risk.

Example: p = 0.4471, there’s a 44.71% chance of an attack per session.

- Expected attacks in next 100 sessions: 44.7

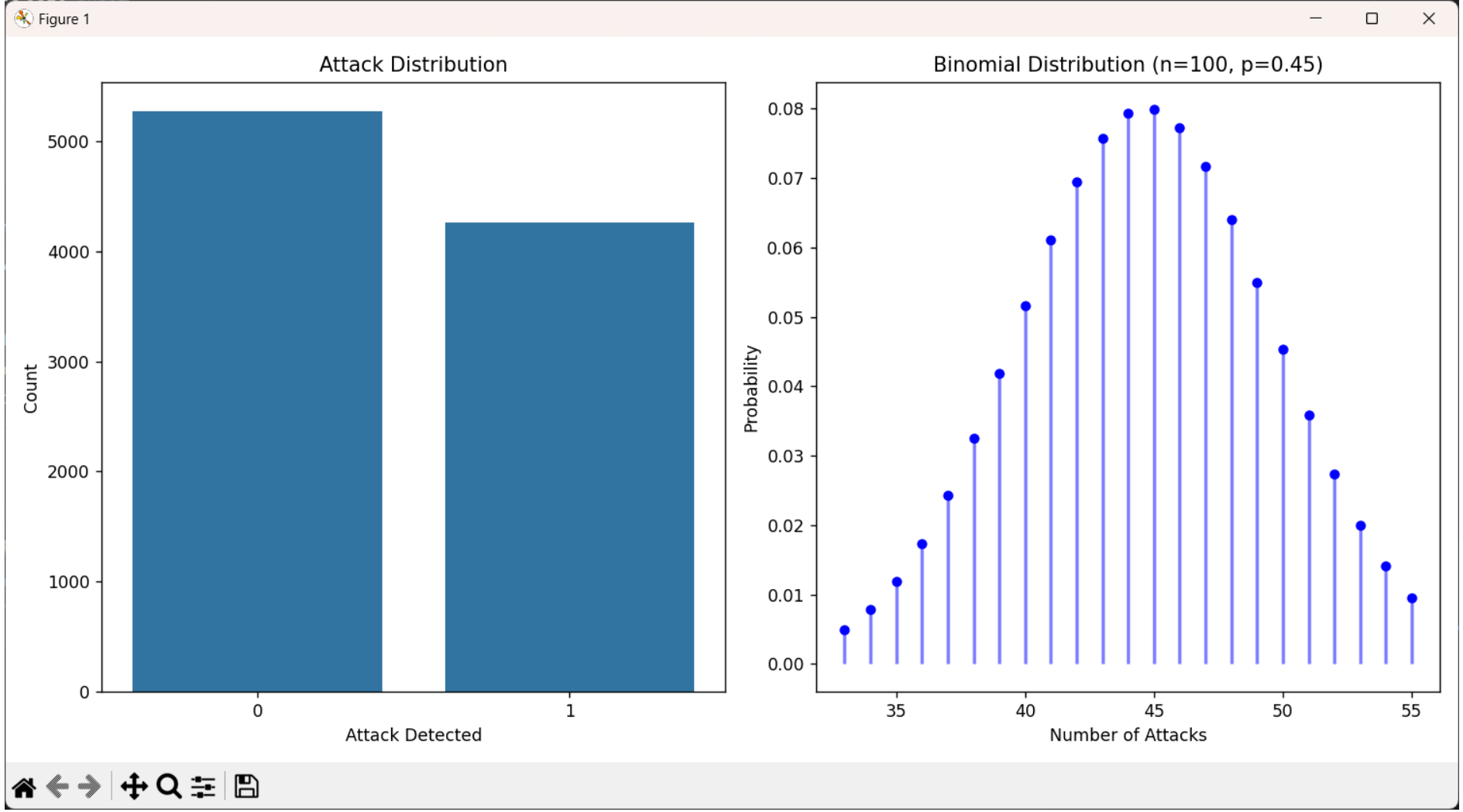
Key findings

protocol_type` : UDP/ICMP has higher attack rates.

- `failed_logins` : Sessions with more failed login attempts may indicate brute-force attacks.

- `encryption_used` : Weak encryption (e.g., DES) could correlate with higher attack success.

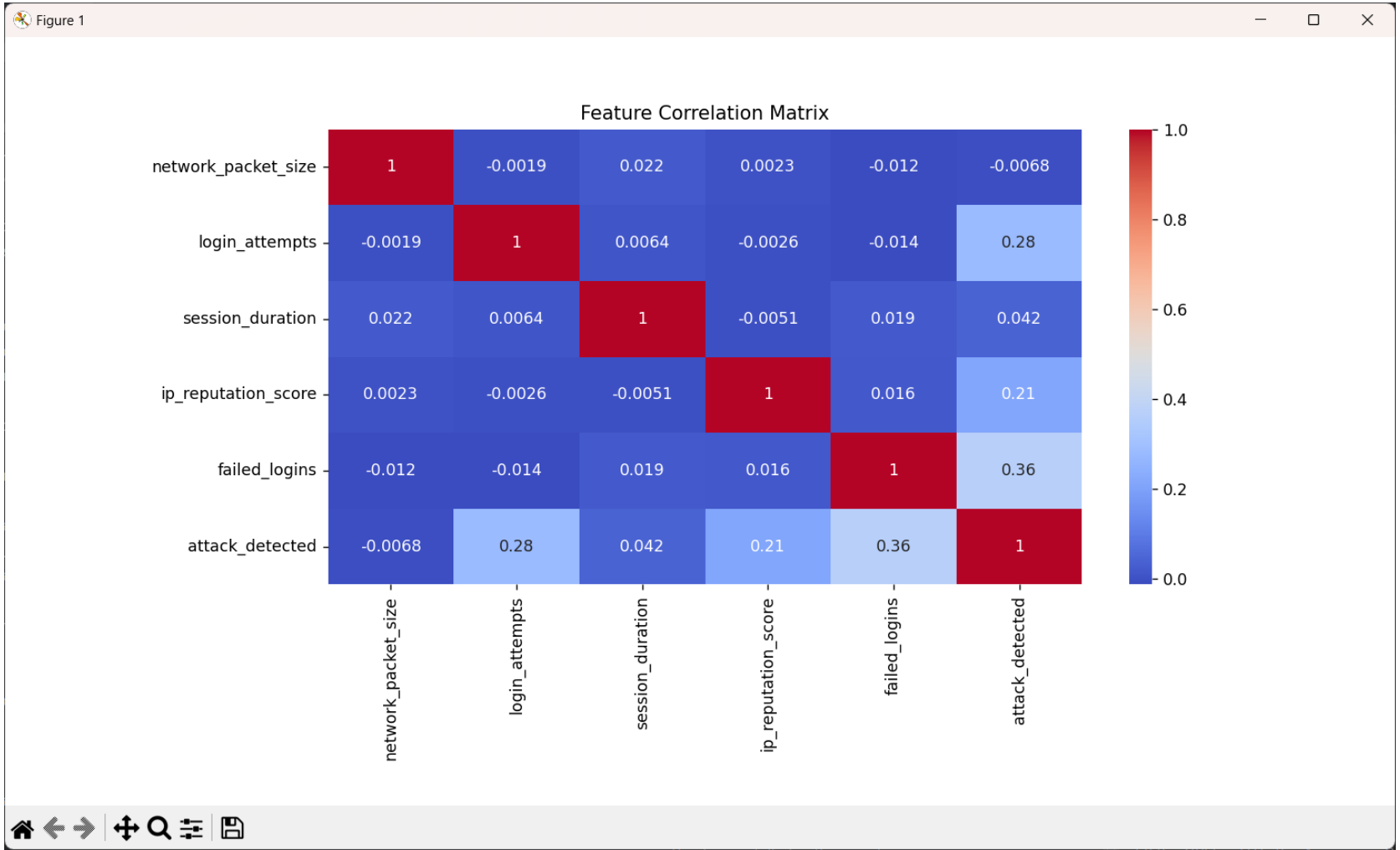
- `ip_reputation_score` : Lower scores may indicate malicious IPs.



1.1 Reveals the **baseline attack rate** in the dataset (e.g., 30% attacks vs. 70% normal traffic)
1.2 Predicts the **likelihood of specific attack counts** (e.g., "There's a 20% chance of 25 attacks in 100 sessions").
1.2 Shows the **spread of risk** (mean ± standard deviation) around expected attacks

. Identifies **risk factors** (e.g., high failed_logins may correlate with attacks).

- Highlights **protective factors** (e.g., high ip_reputation_score may reduce attack likelihood).
- Red = positive correlation, Blue = negative correlation.



Risk Mitigation Measures

- - Enhance Encryption
- - Monitor Suspicious Activity
- - Block Malicious IPs
- - Limit High-Risk Protocols
- - Implement Account Lockout Policies

Normal Distribution Analysis

- Assumptions:
- - Independence
- - No outliers
- - Sufficient sample size

- Statistical Check:
- - Shapiro-Wilk Test: p-value > 0.05 indicates normality.

Normal Distribution Results

- Mean: 792.75
- Standard Deviation: 786.56
- Shapiro-Wilk p-value: 0.0000 (Data is NOT normally distributed)

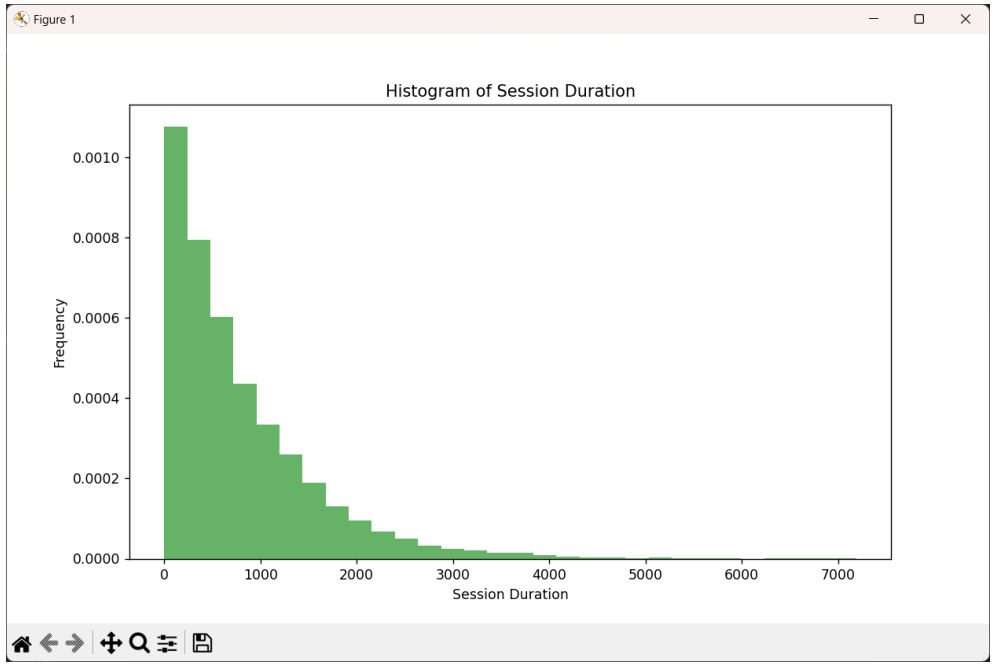


Figure 1
Histogram of Session Duration
Frequency
Session Duration
0.0000
0.0002
0.0004
0.0006
0.0008
0.0010
0
1000
2000
3000
4000
5000
6000
7000

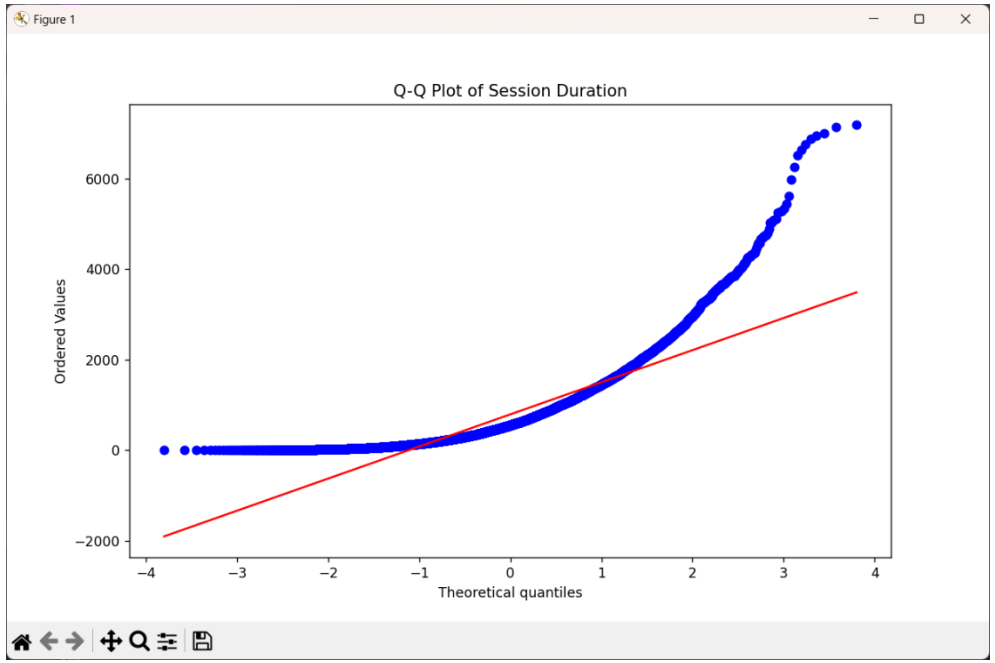


Figure 1
Q-Q Plot of Session Duration
Ordered Values
Theoretical quantiles
−2000
0
2000
4000
6000
−4
−3
−2
−1
0
1
2
3
4

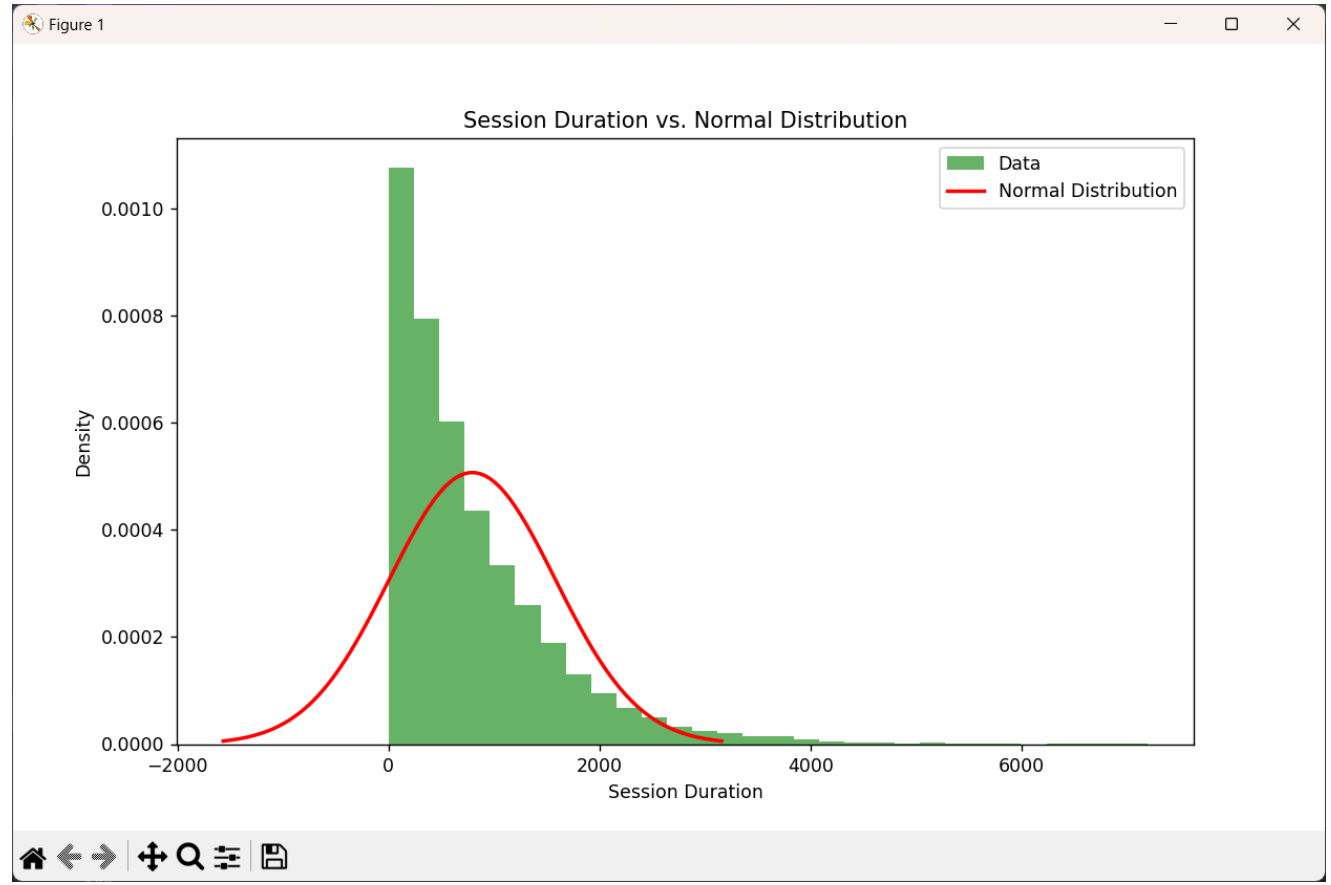


Figure 1
Session Duration vs. Normal Distribution
Data
Normal Distribution
Density
Session Duration
0.0000
0.0002
0.0004
0.0006
0.0008
0.0010
−2000
0
2000
4000
6000

Weibull Distribution Analysis

- Used to model time-to-event (attack detection time).
- **Shape Parameter (β)**:
- - $\beta < 1$: Decreasing hazard rate
- - $\beta \approx 1$: Constant hazard rate
- - $\beta > 1$: Increasing hazard rate
- **Scale parameter (α - Alpha):**This tells us when **63.2% of attacks** have happened

Preparing the Data

- **session_duration:** How long a session lasts before an attack happens or it ends normally.
- **attack_detected:** Whether an attack was found (1 = yes, 0 = no).
- **Censoring:** If no attack happens in a session, we treat it as "incomplete data" (right-censored).

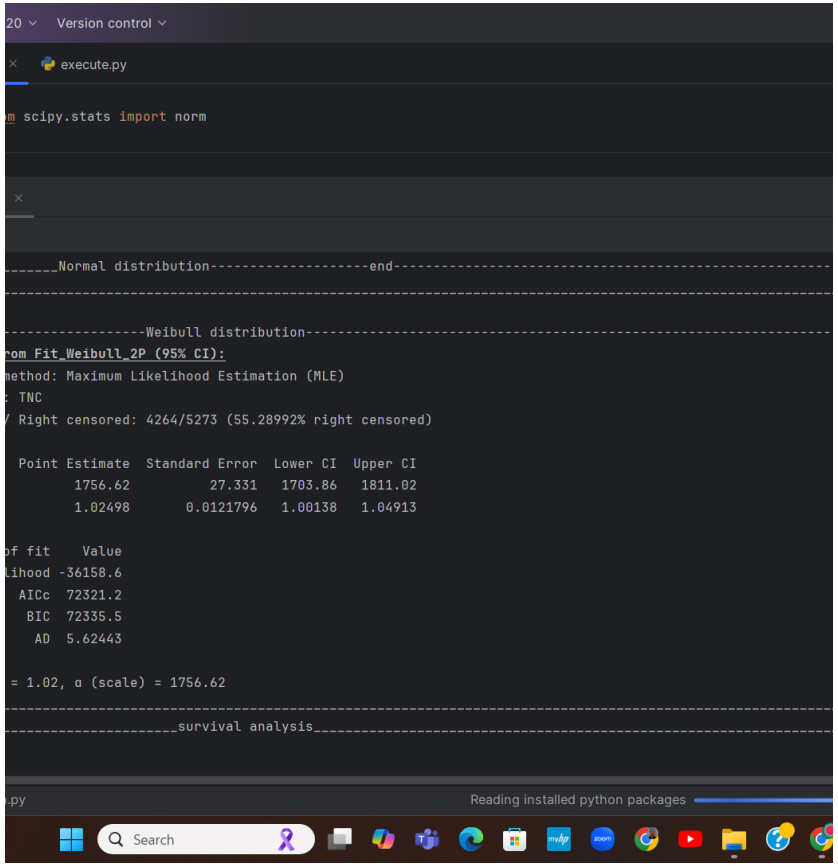
Results

- 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.

- **Right-Censored Data (55.29%)-**55.29%

of the sessions **ended without an attack**

(censored data).



Possible Security Strategies – Expert Opinion

General Continuous Monitoring- Rsk is constant so attack can occur at any time.

Authentication & Initial Security Measures: Focus on **strong authentication at the start of a session** to prevent unauthorized access early.

Session Timeout Consideration: Unlike a scenario where attacks increase over time ($\beta > 1$), setting a hard session limit (like 600 seconds) may not significantly reduce risk in this case.

Survival Analysis

Survival analysis is used to estimate how long something will last before an event happens. In this case, we are analyzing **how long a session lasts before an attack happens** using the Kaplan-Meier estimator.

Preparing the Data

The data is sorted based on session duration (asc).

Sessions where an attack happens are labeled as **events**.

Sessions that end without an attack are **right-censored.**

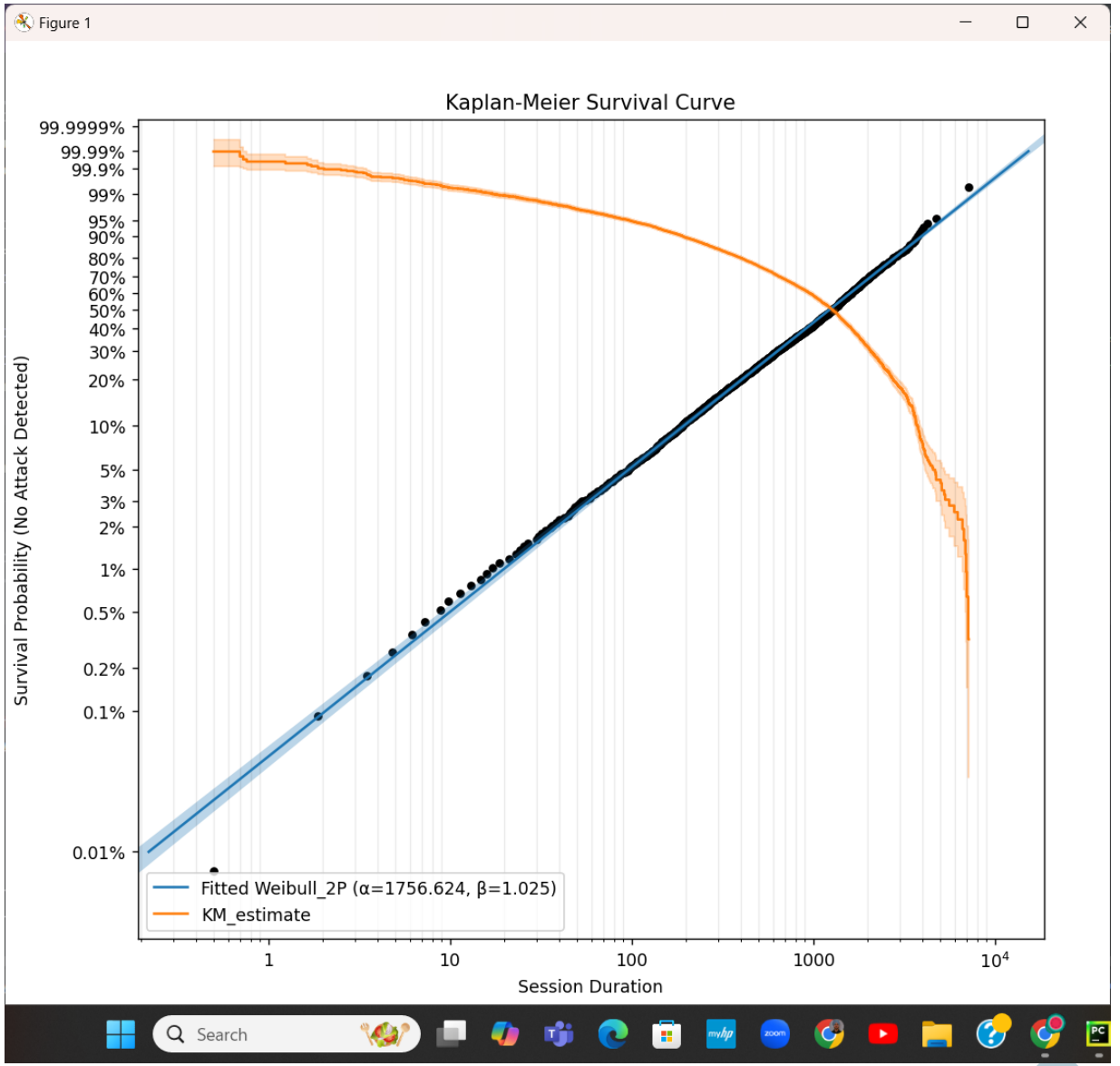


Figure 1
Kaplan-Meier Survival Curve
Survival Probability (No Attack Detected)
99.9999%
99.99%
99.9%
99%
95%
90%
80%
70%
60%
50%
40%
30%
20%
10%
5%
3%
2%
1%
0.5%
0.2%
0.1%
0.01%
1
10
100
1000
10^4
Session Duration
Fitted Weibull_2P (α=1756.624, β=1.025)
KM_estimate
Search

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

Review and Outcome

- We were able to perform a variety of Risk Analysis on varying features of our dataset.
- Our study helped understand how various distributions work on specific features of the dataset.