# ZERO DAY ATTACK DETECTION ON LOGISTICS NETWORK

**JAYESH PATIL**

Module 6
Final Project Report

25th October, 2024

—

ITC6000 - Database Management Systems

—

Prof. Mohsen Bahrami

# INTRODUCTION

This project centers on a critical application in airport logistics networks, specifically a **Zero-Day Attack Detection* system designed to protect sensitive operations against cyber threats. Cybersecurity is most important in today's interconnected supply chains, especially within logistics networks. This application aims to detect real-time anomalies, aiding Security Analysts, IT Administrators, and Logistics Managers in efficiently identifying and responding to these hidden threats.

**Ref. Dataset Link- (Dataset Engineer. (n.d.). Zero-day attack detection in logistics networks [Data set]. Kaggle. Retrieved October 25, 2024, from https://www.kaggle.com/datasets/datasetengineer/zero-day-attack-detection-in-logistics-networks)*

App's Cost Model-

The application follows a **subscription-based cost model**, with various tiers based on network size and required support levels. This approach provides flexibility, allowing different airport sizes to adopt the service affordably.

Basic features like general network monitoring and limited alerts are included at no extra charge, while premium subscriptions offer extensive threat detection and deeper analytical reports.

Personally, I am drawn to this project as it aligns with my past academic pursuits in the supply chain, and currently by studying Analytics, by using its tools (SQLite) would like to connect these two aspects and extract and learn meaningful output and explore more into it.

# BUSINESS ANALYSIS

By defining user personas, such as Security Analysts and IT Administrators, and identifying the business rules guiding threat detection, this analysis ensures that the system's design aligns with efficient monitoring, compliance, and response to cyber threats.

User Personas:

1. **Security Analyst**

The Security Analyst is **responsible for monitoring the logistics network** for potential threats and suspicious activities. Utilizing real-time threat detection, the analyst receives alerts on anomalies and can analyze into event descriptions to assess the nature and seriousness of threats. With access to detailed reports and analytics, the Security Analyst can prioritize responses based on risk levels, ensuring quick and effective threat mitigation to safeguard the logistics network from attacks.

2. **IT Administrator**

The IT Administrator **manages the network's technical infrastructure**, ensuring security and stability. This user configures network security protocols, monitors system health, and responds to any detected threats. With tools to track network events, it helps to maintain uninterrupted logistics operations.

3. **Logistics Operations Manager**

Focused on **maintaining logistics continuity**, the Operations Manager monitors how potential threats might impact the movement of goods. By reviewing high-level security summaries and anomaly alerts, they coordinate responses with the security and IT teams. This role ensures logistics flow smoothly, which is crucial in a highly interconnected airport environment.

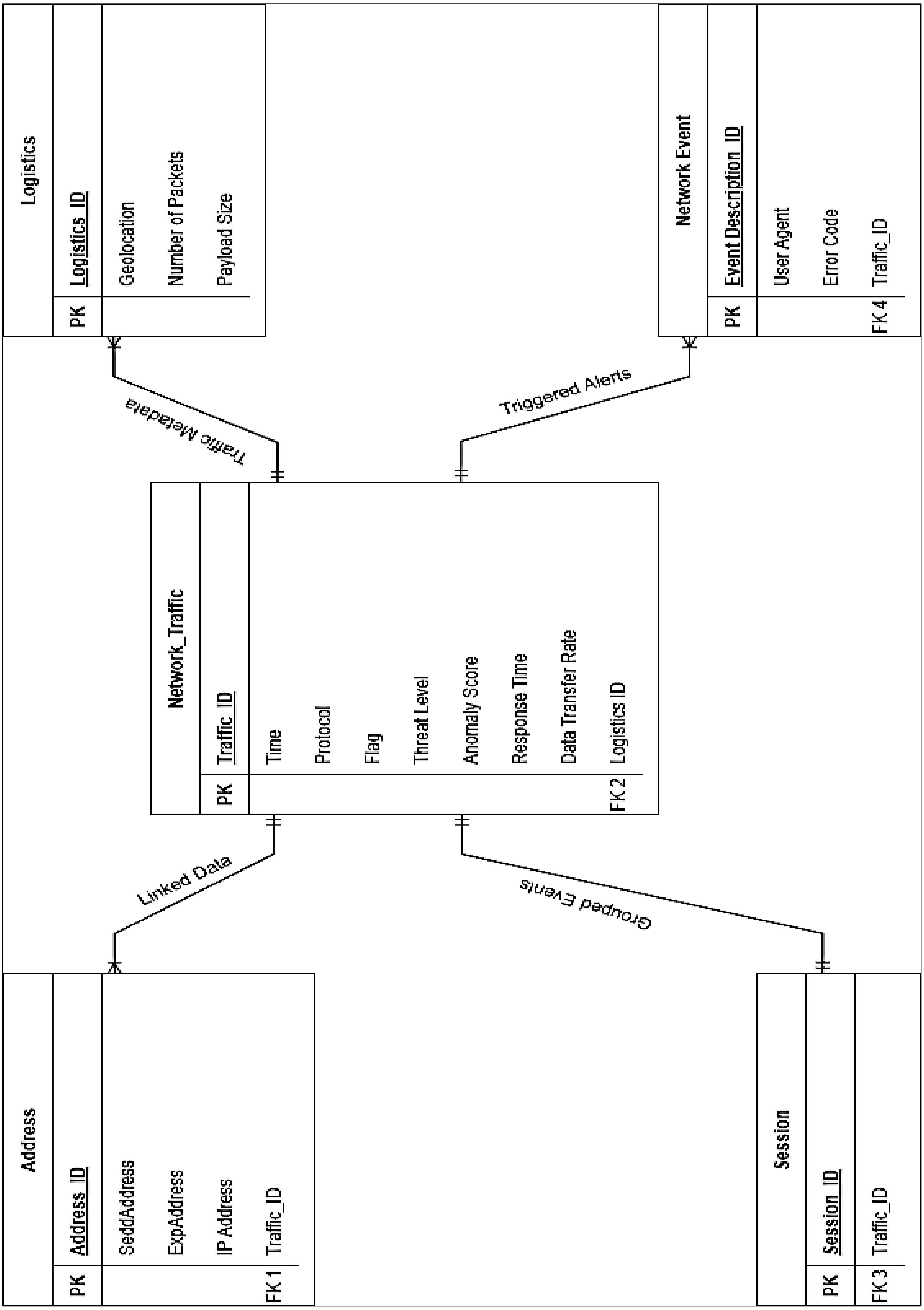
4. **Compliance Officer**

The Compliance Officer ensures that all aspects of network operations **adhere to regulatory standards**, such as GDPR. This role involves reviewing audit logs and reports to verify compliance, monitoring for any breaches, and coordinating with security teams to maintain data integrity.

## BUSINESS RULES/LOGIC

- **Threat Detection**:
  - Continuously monitors incoming network traffic for errors
  - It will generate alerts for any suspicious activities
- **Anomaly Scoring and Risk Assessment**:
  - It will score each detected anomaly based on its threat level.
  - It will categorize threats to prioritize responses.
- **Response**:
  - It will use its previous records to respond protocols automatically when certain threats are detected.
- **Data Storage**:
  - Network logs are stored securely for real-time and historical analysis.
  - Role-based access control ensures data integrity.
- **Reporting and Analytics**:
  - They will generate detailed reports to support decision-making.
  - They will also provide high-level and detailed insights based on user roles.

# TABLE DESIGN & ANALYSIS



*diagrams.net. (n.d.). diagrams.net (draw.io). Retrieved October 25, 2024, from* https://app.diagrams.net/

## Analysis

The ER diagram you provided includes several key entities that form the structure of your Zero-Day Attack Detection system.

1. **Logistics**
    - Attributes: Logistics_ID (PK), Geolocation, Number of Packets, Payload Size
    - Description: This table stores data of location of traffic, the number of packets being transferred, and the size of each packet.
    - Relationships: One-to-Many with Network_Traffic

    "Each logistics record can be linked to multiple traffic entries."

2. **Network_Traffic**
    - Attributes: Traffic_ID (PK), Time, Protocol, Flag, Threat Level, Anomaly Score, Response Time, Data Transfer Rate, FK_2 Logistics_ID
    - Description: This table stores network traffic information, including protocols, threat levels, and others.
    - Relationships: One-to-Many with Logistics

    "Each traffic entry is linked to a specific logistics record."

        - One-to-Many with Network_Event, Session, and Address

    "This table links to multiple entries in all the tables."

3. **Network_Event**
    - Attributes: Event_Description_ID (PK), User Agent, Error Code, FK_4 Traffic_ID
    - Description: This table captures events related to network traffic, including user agents and error codes.
    - Relationships: One-to-Many with Network_Traffic

    "Each traffic entry can generate multiple events."

4. **Address**
    - Attributes: Address_ID (PK), SendAddress, ExpAddress, IPAddress, FK_1 Traffic_ID
    - Description: This table stores address information, including sending and receiving addresses, and the IP address.
    - Relationships: One-to-Many with Network_Traffic

    "Each traffic entry can link to multiple address records."

### 5. Session

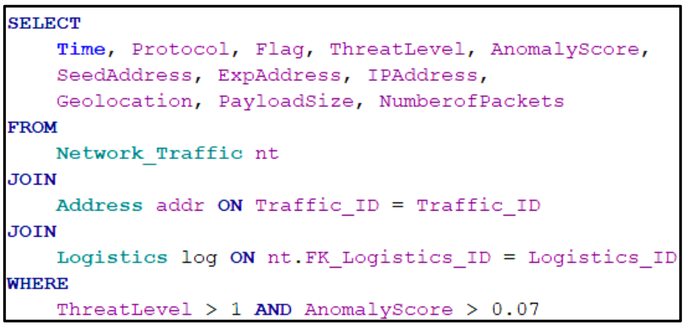
- Attributes: Session_ID (PK), FK_3 Traffic_ID
- Description: This table stores session details that group related traffic events.
- Relationships: One-to-Many with Network_Traffic

  "Each session can consist of multiple traffic entries."

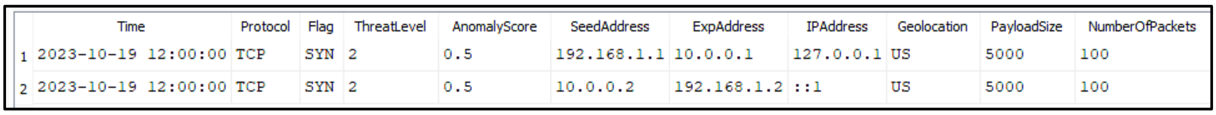
## DATABASE IMPLEMENTATION

*(Have used SQLite app to solve these queries SQLite. (2024, September 25). Release 3.47.0 . Retrieved October 25, 2024, from [https://www.sqlite.org/releaselog/3_47_0.html](https://www.sqlite.org/releaselog/3_47_0.html))*

**Q**. "Provide SQL commands to retrieve network traffic data with high Threat Level and Anomaly Score, joining relevant tables for a comprehensive view of **suspicious activities**."



```
SELECT
    Time, Protocol, Flag, ThreatLevel, AnomalyScore,
    SeedAddress, ExpAddress, IPAddress,
    Geolocation, PayloadSize, NumberofPackets
FROM
    Network_Traffic nt
JOIN
    Address addr ON Traffic_ID = Traffic_ID
JOIN
    Logistics log ON nt.FK_Logistics_ID = Logistics_ID
WHERE
    ThreatLevel > 1 AND AnomalyScore > 0.07
```



| | Time | Protocol | Flag | ThreatLevel | AnomalyScore | SeedAddress | ExpAddress | IPAddress | Geolocation | PayloadSize | NumberOfPackets |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2023-10-19 12:00:00 | TCP | SYN | 2 | 0.5 | 192.168.1.1 | 10.0.0.1 | 127.0.0.1 | US | 5000 | 100 |
| 2 | 2023-10-19 12:00:00 | TCP | SYN | 2 | 0.5 | 10.0.0.2 | 192.168.1.2 | ::1 | US | 5000 | 100 |

Interpretation- The table displays two network events flagged with a ThreatLevel of 2 and an AnomalyScore of 0.5, both using the TCP protocol with SYN flags, locating from IP addresses within the US. Each event has a PayloadSize of 5000 and involves 100 packets, suggesting consistent, potentially suspicious traffic activity.

Story and Use Case for Suspicious Activity-

**Scenario:**

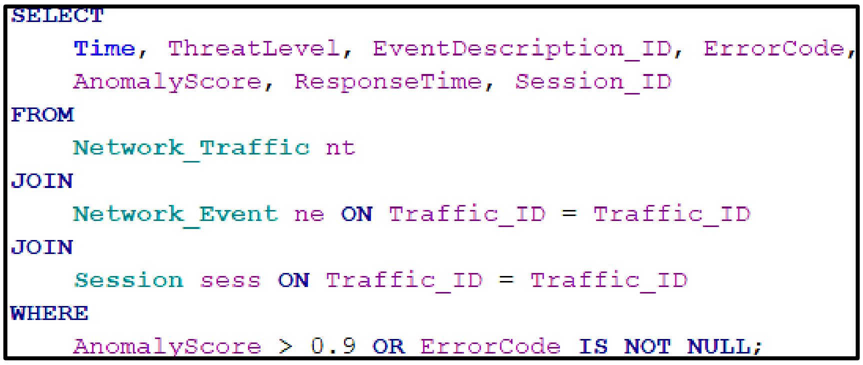
A security analyst needs to retrieve network events with high threat levels and anomaly scores, indicating potential risks.

**User Registers Suspicious Activity:**
The system detects higher rate in anomaly scores and registers high-risk events, joining data from Network_Traffic, Address, and Logistics tables for a full view of network activity.

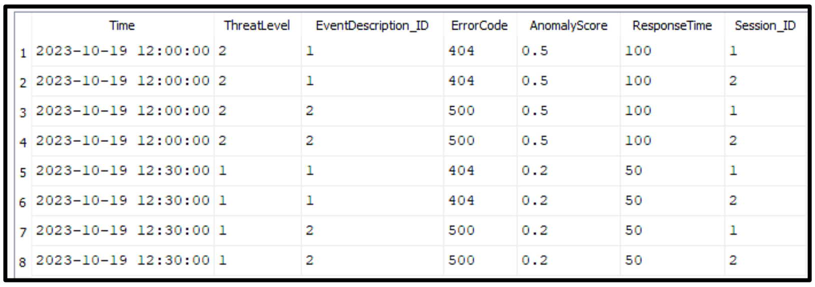
**Investigation query:**
The analyst uses this query to retrieve events with high ThreatLevel and AnomalyScore, accessing detailed information on traffic, IP, geolocation, and packet details to identify suspicious activity.

**Action:**
The analyst identifies harmful IPs or geolocations and takes action, such as blocking. This query is crucial for real-time threat detection and response.

**Q**. "Provide SQL commands to retrieve network traffic data by joining multiple tables, focusing on events with high AnomalyScore or ErrorCode for **monitoring and diagnosing network issues."**



```
SELECT
    Time, ThreatLevel, EventDescription_ID, ErrorCode,
    AnomalyScore, ResponseTime, Session_ID
FROM
    Network_Traffic nt
JOIN
    Network_Event ne ON Traffic_ID = Traffic_ID
JOIN
    Session sess ON Traffic_ID = Traffic_ID
WHERE
    AnomalyScore > 0.9 OR ErrorCode IS NOT NULL;
```



| | Time | ThreatLevel | EventDescription_ID | ErrorCode | AnomalyScore | ResponseTime | Session_ID |
|---|---|---|---|---|---|---|---|
| 1 | 2023-10-19 12:00:00 | 2 | 1 | 404 | 0.5 | 100 | 1 |
| 2 | 2023-10-19 12:00:00 | 2 | 1 | 404 | 0.5 | 100 | 2 |
| 3 | 2023-10-19 12:00:00 | 2 | 2 | 500 | 0.5 | 100 | 1 |
| 4 | 2023-10-19 12:00:00 | 2 | 2 | 500 | 0.5 | 100 | 2 |
| 5 | 2023-10-19 12:30:00 | 1 | 1 | 404 | 0.2 | 50 | 1 |
| 6 | 2023-10-19 12:30:00 | 1 | 1 | 404 | 0.2 | 50 | 2 |
| 7 | 2023-10-19 12:30:00 | 1 | 2 | 500 | 0.2 | 50 | 1 |
| 8 | 2023-10-19 12:30:00 | 1 | 2 | 500 | 0.2 | 50 | 2 |

**Interpretation-** The table shows multiple network events occurring at different times, with varying ThreatLevel, EventDescription_ID, ErrorCode, AnomalyScore, and ResponseTime. Events with ErrorCode 404 and 500 have moderate AnomalyScores (0.5 and 0.2) and are grouped into sessions, which few are repeated, which signifies persistent issues that need investigation.

<u>Story and Use Case for monitoring and diagnosing network issues</u>

- **Scenario:**
  A security analyst needs insights into events with high anomaly scores or specific error codes, such as 404 (not found) or 500 (server error), which might indicate repeated issues within certain network sessions.

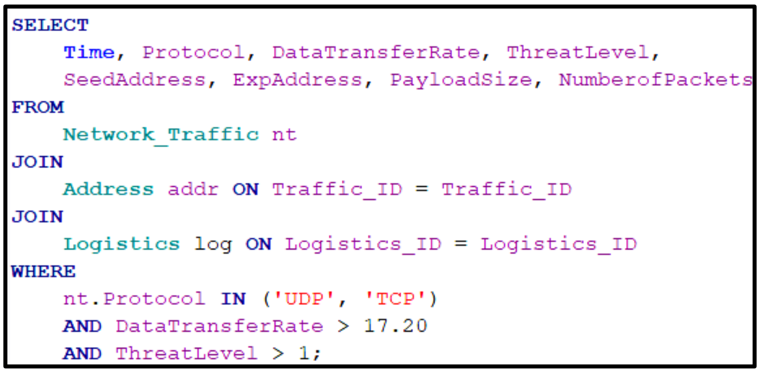
- **User Registers Error Activity:**
  When the system detects events with significant error codes or high anomaly scores, it logs these events, linking data from Network_Traffic, Network_Event, and Session tables to capture a detailed view of the problematic network activities.

- **Investigation Query:**
  The analyst uses this query to pull all relevant events where the AnomalyScore is high

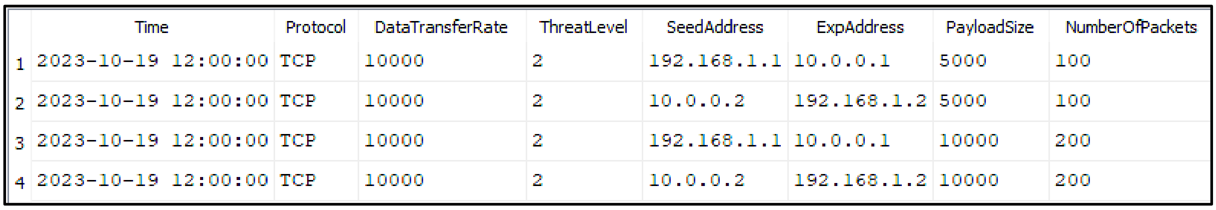
The output provides essential details like ThreatLevel, ErrorCode, ResponseTime, and Session_ID, helping the analyst to quickly identify problematic sessions and errors.

- **Action:**
  Based on the query results, the analyst can identify specific sessions and IP addresses associated with repeated errors and take action to investigate the cause, such as checking for errors.

**Q**. "Provide SQL commands to **retrieve high-risk network traffic data** for UDP and TCP protocols with high data transfer rates, joining multiple tables for detailed analysis."



```
SELECT
    Time, Protocol, DataTransferRate, ThreatLevel,
    SeedAddress, ExpAddress, PayloadSize, NumberofPackets
FROM
    Network_Traffic nt
JOIN
    Address addr ON Traffic_ID = Traffic_ID
JOIN
    Logistics log ON Logistics_ID = Logistics_ID
WHERE
    nt.Protocol IN ('UDP', 'TCP')
    AND DataTransferRate > 17.20
    AND ThreatLevel > 1;
```



| | Time | Protocol | DataTransferRate | ThreatLevel | SeedAddress | ExpAddress | PayloadSize | NumberOfPackets |
|---|---|---|---|---|---|---|---|---|
| 1 | 2023-10-19 12:00:00 | TCP | 10000 | 2 | 192.168.1.1 | 10.0.0.1 | 5000 | 100 |
| 2 | 2023-10-19 12:00:00 | TCP | 10000 | 2 | 10.0.0.2 | 192.168.1.2 | 5000 | 100 |
| 3 | 2023-10-19 12:00:00 | TCP | 10000 | 2 | 192.168.1.1 | 10.0.0.1 | 10000 | 200 |
| 4 | 2023-10-19 12:00:00 | TCP | 10000 | 2 | 10.0.0.2 | 192.168.1.2 | 10000 | 200 |

Interpretations- The table displays high-risk TCP events with a data transfer rate of 10,000 and threat level of 2, showing alternating source and destination addresses, large payloads, and high packet counts, indicating potentially suspicious data transfers.

Story and Use Case for retrieving high-risk network traffic data

- **Scenario:**
  A security analyst focuses on events where DataTransferRate is high and ThreatLevel is above 1, needing details on source/destination addresses, protocols, payload sizes, and packet counts.

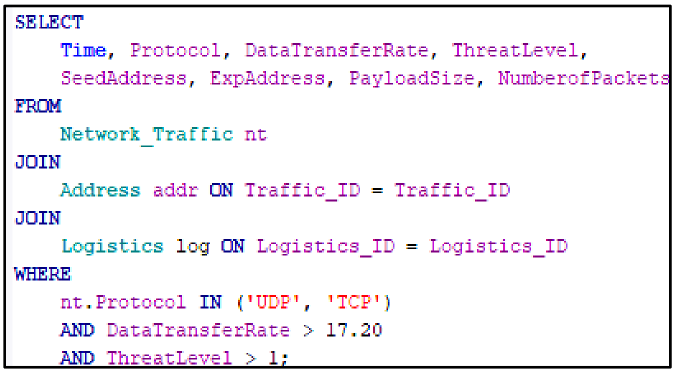
- **User Registers High-Risk Traffic Activity:**
  When the system detects a DataTransferRate over 17.20 on specified protocols, it logs these events by joining data from Network_Traffic, Address, and Logistics tables, creating a record for further analysis.

- **Investigation Query:**
  The analyst retrieves events with high data transfer rates and threat levels, getting insights into addresses, protocol types, payload size, and packet details to evaluate the risk level.

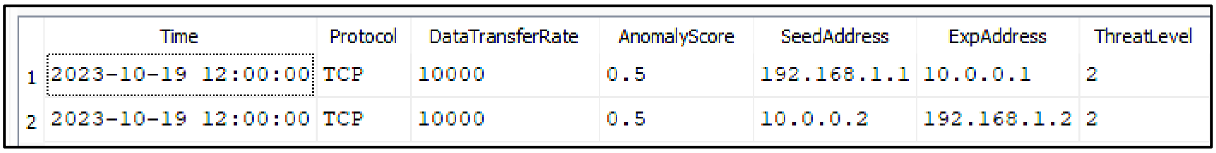
- **Action:**
  The analyst examines high-risk events and takes action on suspicious IPs, potentially restricting or monitoring traffic to prevent data breaches or network issues.

## ANALYTICS, REPORTS & METRICS

**Q**. Find TCP sessions with unusually high data transfer rates to monitor for potential suspicious activity.



```
SELECT
    Time, Protocol, DataTransferRate, ThreatLevel,
    SeedAddress, ExpAddress, PayloadSize, NumberofPackets
FROM
    Network_Traffic nt
JOIN
    Address addr ON Traffic_ID = Traffic_ID
JOIN
    Logistics log ON Logistics_ID = Logistics_ID
WHERE
    nt.Protocol IN ('UDP', 'TCP')
    AND DataTransferRate > 17.20
    AND ThreatLevel > 1;
```

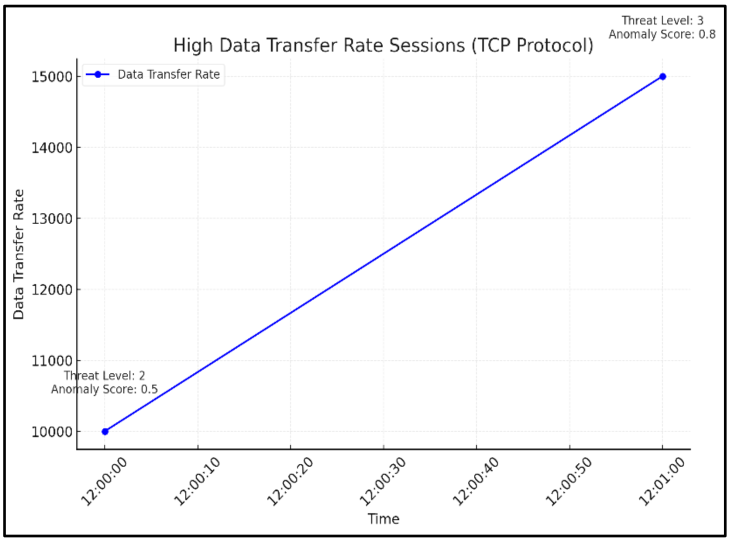


| | Time | Protocol | DataTransferRate | AnomalyScore | SeedAddress | ExpAddress | ThreatLevel |
|---|---|---|---|---|---|---|---|
| 1 | 2023-10-19 12:00:00 | TCP | 10000 | 0.5 | 192.168.1.1 | 10.0.0.1 | 2 |
| 2 | 2023-10-19 12:00:00 | TCP | 10000 | 0.5 | 10.0.0.2 | 192.168.1.2 | 2 |

Interpretation-This report shows TCP sessions with very high data transfer rates, at 10,000, which are well above standard levels. Although the threat levels are low (2), the high transfer rate and anomaly score (0.5) suggest unusual activity.

By focusing on these sessions, security analysts can investigate potential risks, as unusually high data transfer rates can sometimes indicate data leaks.

Line chart –High Data Transfer Rate Sessions (TCP Protocol)



**I created the graph using **Matplotlib**, from Python library.*

This line graph represents high data transfer rate sessions over time for the TCP protocol:

- **X-axis:** Time of the session.
- **Y-axis:** Data Transfer Rate.

Interpretation- The data shows two high data transfer sessions on October 19. The first session at 12:00 has a data transfer rate of 10,000 units with a Threat Level of 2 and an Anomaly Score of 0.5.

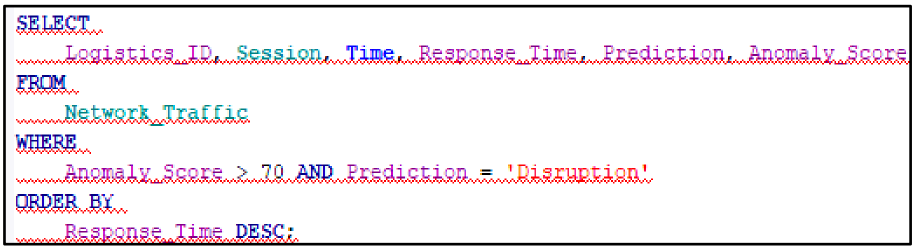
The second session at 12:01 has an even higher transfer rate of 15,000 units, with a Threat Level of 3 and an Anomaly Score of 0.8, indicating increased risk.

This upward trend in transfer rate, with the rising threat level, suggests a need for closer monitoring.

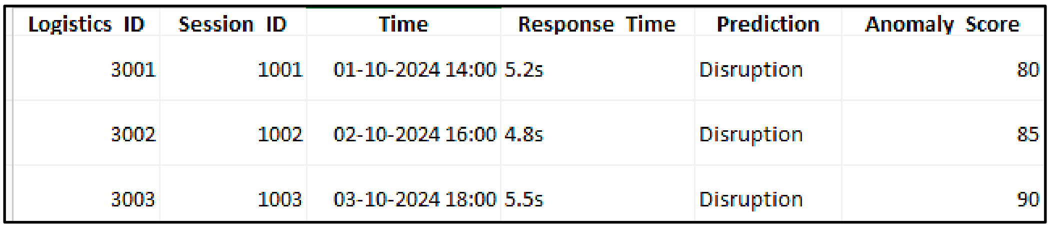
**Q**. Identify network sessions with high response times predicted to disrupt logistics operations.

**Explanation**

This query aims to detect network events that could impact logistics due to high response times and disruption predictions.



```
SELECT
    Logistics_ID, Session, Time, Response_Time, Prediction, Anomaly_Score
FROM
    Network_Traffic
WHERE
    Anomaly_Score > 70 AND Prediction = 'Disruption'
ORDER BY
    Response_Time DESC;
```



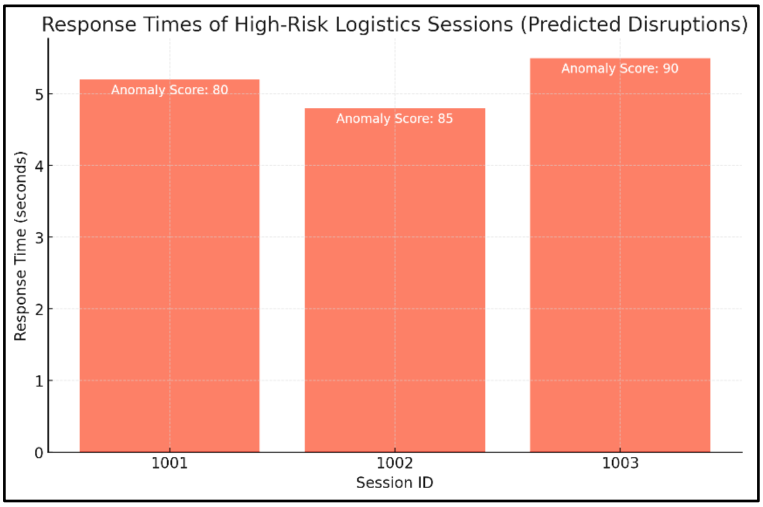
| Logistics ID | Session ID | Time | Response Time | Prediction | Anomaly Score |
|---|---|---|---|---|---|
| 3001 | 1001 | 01-10-2024 14:00 | 5.2s | Disruption | 80 |
| 3002 | 1002 | 02-10-2024 16:00 | 4.8s | Disruption | 85 |
| 3003 | 1003 | 03-10-2024 18:00 | 5.5s | Disruption | 90 |

**Interpretation**

Sessions with an anomaly score above 70 indicate deviations from typical network performance, and predictions labeled as "Disruption" suggest that these sessions have a high risk of causing operational delays.

By focusing on these sessions, logistics managers can take proactive measures to avoid disruptions.

Bar Graph- Response Times of High-Risk Logistics Sessions (Predicted Disruptions)



**I created the graph using **Matplotlib**, from Python library.*

**Interpretation**

The bar chart illustrates the response times of network sessions predicted to disrupt logistics operations. Each bar represents a session, with the height indicating the Response Time in seconds.

- **Session 1003** has the highest Response Time of 5.5 seconds and an Anomaly Score of 90, indicating it is the most significant disruption risk.
- **Session 1002** follows with a Response Time of 4.8 seconds and an Anomaly Score of 85.
- **Session 1001** has a Response Time of 5.2 seconds and an Anomaly Score of 80.

This visualization helps prioritize which sessions to investigate first, focusing on those with high response times and elevated anomaly scores that may disrupt logistics.

## SECURITY CONCERNS

*(Ground Labs. (n.d.). Data security in transportation and logistics: Protecting sensitive information in a growing industry. Retrieved October 25, 2024, from https://www.groundlabs.com/blog/data-security-transportation-logistics/)*

The Zero-Day Attack Detection system processes and stores various types of sensitive information critical to ensuring network security and operational continuity.

As I being the data expert of this app, I am aware that every information, each carrying specific privacy and security risks that other stakeholders (e.g., security team, legal, privacy departments) should be informed about: Below are the key security and privacy concerns associated with this data:

1. **Sensitive Data Storage**

    ○ **Network Traffic Logs**: These logs capture detailed network activity, such as IP addresses, protocols, data transfer rates, and anomaly scores. People who are not allowed to access to this data could expose network behavior patterns, and would be alerted immediately.

  - **User Access Information**: This includes data on user roles, permissions, and access logs, which is sensitive as it reveals who has access to critical system functions. Ensuring RBAC (role-based access control) and monitoring is essential to prevent unauthorized access.
  - **Logistics Data**: Logistics data is vital for tracking and securing supply chain operations. Unauthorized access to this information could disrupt logistics operations.
  - **Incident Reports**: These reports document security incidents, detected anomalies, and response actions. If accessed by unauthorized parties, they could reveal weaknesses in the system's defenses, providing attackers a chance to hack into our system.

2. **High-Value Targets for Hackers**
   - **Network Traffic Logs**: Attackers may target these logs to understand network flows and detect weak points.
   - **User Access Information**: By gaining access to user access details, attackers can conduct phishing or social engineering attacks to compromise users.
   - **Logistics Data**: Due to its importance in supply chain continuity, logistics data is a high-value target. Compromising this data could lead to disruptions or unauthorized manipulations in logistics.
3. **Key Security Recommendations**
   To protect this sensitive information, the data expert recommends the following measures:
   - **Data Encryption**: All sensitive data, especially network logs, user access information, and logistics details, should be encrypted both in transit and at rest to prevent unauthorized access.
   - **Access Control and Role Management**: Implementing strict access control mechanisms, such as role-based access control (RBAC), ensures that only authorized personnel can access sensitive data.
   - **Compliance and Logging**: Adhering to regulatory standards, such as GDPR, requires robust logging policies. All access to sensitive data should be logged in the system.

These security measures help mitigate the risks associated with processing and storing sensitive information, safeguarding the organization.

# ARCHITECTURE

Suggested Solution Architecture

The Zero-Day Attack Detection system uses a **three-tier client-server model** in a **hybrid cloud environment** to provide scalability, security, and efficient data processing.

1. **Three-Tier Architecture**
    - **Client Tier**: The client tier includes a web and mobile dashboard built with React.js, accessible through HTTP/HTTPS APIs. This dashboard allows them to monitor network traffic, detect threats, and respond promptly from any device.
    - **Application Tier**:
        - **Anomaly Detection**: It will use machine learning models, such as Random Forest and Deep Learning, to identify unusual patterns in network traffic.
        - **Traffic Monitoring**: By using Python it will process real-time data in enabling immediate detection.
    - **Data Tier**: The data layer consists of a relational database for structured data storage and MongoDB for handling high-volume traffic logs. This setup is good at retrieval and storage of both structured and unstructured data.
2. **Hosting Model: Hybrid Cloud**
    - **Cloud Hosting** (AWS, Azure): Used for scalable, long-term storage, backup, and disaster recovery, supporting the system's need for flexible storage.
    - **On-Premise Hosting**: Essential for processing highly sensitive data in real-time, ensuring control over critical data.
3. **Security Architecture**
    - **Network Security**: It will implements Intrusion Detection/Prevention Systems (IDS/IPS), VPN, encryption to secure data communication.
    - **Data Security**: Uses role-based access control (RBAC) and multi-factor authentication (MFA) to restrict data access to authorized users.
    - **Auditing and Monitoring**: Leveraging tools like "AWS CloudTrail" and to maintain an audit trail of system actions for suspicious activity.
4. **Storage Requirements**
    - **Initial Storage Estimate**: 100 GB to 1 TB to accommodate daily logs and traffic data.
    - **Long-Term Storage**: For larger enterprises, up to 10 TB to store 6-12 months of data for historical analysis and compliance purposes.

- **Cloud Backup**: On-demand, scalable storage to handle increased data retention needs without compromising performance.

## PROJECT WRAP-UP & FUTURE CONSIDERATIONS

This project has taught me a lot about building secure systems to protect against cyber threats, especially zero-day attacks in logistics networks. Here are some key takeaways:

1. Working with sensitive data like network and user info showed me the importance of strong security measures like encryption and access controls. I'll prioritize data protection in future projects, especially for sensitive information.
2. Using both cloud and on-premise storage gave a good knowledge. Cloud storage is easy to scale, while on-premise keeps important data safe—an approach I'll use in the future for secure, scalable storage.
3. Learned how to create an ER diagram, including notations and setting up relationships between tables, which would be beneficial for me in the future in simplifying the data analyzing it better, and delivering meaningful insights.
4. Gained an understanding of the backend operations needed for a functional database architecture, which will help in my career ahead
5. Developed insights into securing databases and protecting sensitive information on the further projects I will be working on.
6. Learned from Module 6 discussions on how data breaches occur and the steps needed to prevent them.

Overall, I have learned many new things in terms of SQLite tool, understanding databases, making tables, finding out alternatives to join attributes and solving problems, or understanding what we specifically need.

This project and course have given me an added advantage in my skillset and in my CV for my professional career ahead.

Have Learned many lessons on this academic journey.