Anomaly Types (Figure 9 from the original paper)

The original paper identifies several types of anomalies, as visualized in **Figure 9**:

1. **Point Anomaly - Global**: Represents a sharp and isolated deviation from the overall pattern across the entire dataset.

2. **Point Anomaly - Contextual**: A data point that deviates from the normal context in a specific time window.

3. **Collective Anomaly - Subsequence**: A pattern anomaly where a series of data points together form an unusual sequence, even if individual points don't appear anomalous.

4. **Collective Anomaly - Pattern**: A group of data points forming an unexpected pattern when compared to the general trends of the series.

5. **Seasonal Anomaly**: A seasonal fluctuation in the data that deviates significantly from the expected cyclic behavior.

6. **Trend Anomaly**: A change in the underlying trend of the time series data, such as a sudden shift from a positive to a negative slope.

Time Series Metrics for Anomaly Detection (Figure 10 from the original paper)

The original paper analyzes multiple metrics (like n_flows, n_packets, n_bytes, etc.) for the time series of IP address ID 1367, identified as a **Denial of Service (DoS)** attack. The highlighted anomalies in **Figure 10** correspond to a period where abnormal traffic behaviors were observed, especially around **2024-05-01** to **2024-06-15**. Key observations include:

- A sharp spike in **n_flows** and **n_packets**, but with a relatively smaller increase in **n_bytes**, indicating a DoS attack where many packets were sent without a proportional increase in data volume.

- Significant shifts in metrics like **dir_ratio_bytes** and **avg_duration**, which further corroborate the DoS nature of the anomaly.

- The **avg_ttl** remained fairly consistent, as expected in a DoS attack where traffic originates from a single source.

Comparison with Your Results

Looking at the **Isolation Forest Anomalies** in our figures, we can compare the patterns as follows:

Detected Anomalies in Your Results

- Our figures (e.g., **Batch 34**, **Batch 39**, and **Batch 43**) show anomalies detected using Isolation Forest. Each batch shows a spike in the number of flows (n_flows) and a corresponding detection of anomalous behavior (marked as red dots). These spikes align with the patterns observed in **Figure 9** (e.g., **Point Anomaly** or **Trend Anomaly**), but the detected anomalies may differ slightly in frequency due to the smaller dataset and different parameter choices in Isolation Forest.

Specific Points of Comparison

1. **Point Anomalies**: The spikes detected in our figures resemble the **Point Anomaly - Global** type, where a sharp deviation is observed. In **Batch 34**, for example, there's a noticeable spike that could be linked to a global outlier anomaly. The timing also corresponds with **2024-05** to **2024-06** (similar to the red-highlighted periods in **Figure 10** of the paper).

2. **Seasonal Anomaly**: In our plots, there appears to be a recurring pattern of spikes during specific times, which could resemble the **Seasonal Anomaly** seen in **Figure 9**. This is more evident in longer time series data or when the full dataset is analyzed.

3. **Trend Anomaly**: The sharp shifts in traffic volumes are indicative of a potential **Trend Anomaly**, similar to what's observed in **Figure 10**.

4. **Missing Details**: Since our dataset is a smaller subset, the exact identification of anomalies differ slightly in granularity or the range of anomalies detected, but the general trend of detecting unusual traffic behavior (similar to DoS attacks or network misconfigurations) holds.

Reasons for Differences

- **Dataset Size**: The full dataset, as mentioned in the paper, is large, with over 275,000 IP addresses. Using a subset of the data may lead to differences in anomaly detection frequency.
- **Model Parameters**: The Isolation Forest model we used has different hyperparameters (e.g., number of trees, contamination ratio) compared to those used in the original research. These settings can affect how anomalies are detected.
- **Data Preprocessing**: Our smaller dataset, with specific filtering and sampling choices, may not fully replicate the complex network behaviors captured in the full dataset.