UNIVERSITY of HOUSTON

COLLEGE of NATURAL SCIENCES & MATHEMATICS

Enhancing Network Security Awareness: Visual Cyber Threat Analysis

Lokeshwar Reddy Nandanapalli (2299460) Ben Gideon Dokiburra (2283917)

Abhigna Sowgandhika Vadlamudi (2268166)

PROBLEM DESCRIPTION

- Cybersecurity professionals and researchers often deal with complex datasets related to network intrusions and security incidents.

- The goal of this project is to develop effective visualization techniques for the KDD Cup 1999 Data Computer Security Dataset.

- These datasets contain information about network intrusions and normal activities, and visualization of this data helps obtain valuable insights.

VISUALIZATION TECHNIQUES

- *3D Principal Component Analysis (PCA) plot* - by Lokeshwar Reddy Nandanapalli
- *t-distributed Stochastic Neighbor Embedding (t-SNE)* - by Abhigna Sowgandhika Vadlamudi
- *k-means* to visualize clusters of attack types - by Ben Gideon Dokiburra
- *Bar graph and pie chart* to visualize number of attack types - by Lokeshwar Reddy Nandanapalli and Ben Gideon Dokiburra

PRINCIPAL COMPONENT ANALYSIS (PCA)

- It helps in capturing the underlying structures and patterns within the high-dimensional feature space.

- PCA reduces the dimensionality of the dataset while preserving its variance, allowing for the transformation of complex data into a lower-dimensional representation.

- By plotting the principal components in a three-dimensional space, distinct clusters or patterns in the data become discernible.

- This technique provides a holistic overview of the dataset's intrinsic structure, enabling cybersecurity professionals to identify potential correlations and anomalies

t-DISTRIBUTED STOCHASTIC NEIGHBOUR EMBEDDING (t-SNE)

- Emphasizes the local relationships between data points.

- This technique is particularly useful for revealing intricate structures and clusters within the dataset that might not be apparent in higher-dimensional representations.

- By mapping instances with similar characteristics closer together in the visual space, t-SNE facilitates the identification of distinct groups, shedding light on the underlying nature of network intrusions and normal activities.

K-MEANS CLUSTERING

- It involves grouping data points based on similarity, providing insights into distinct patterns or clusters of network activity.

- This visualization technique is crucial for identifying commonalities among instances of intrusions or normal behavior.

- By leveraging K-Means, the dataset is partitioned into clusters, and each point is assigned to the cluster with the nearest centroid.

- Visualizing these clusters offers a tangible representation of the different types of network intrusions present in the data, aiding in the understanding and classification of malicious activities.

BAR GRAPHS

- They allow visual representation of the distribution of various attack types within the KDD Cup 1999 dataset.

- Each unique attack type is plotted along the x-axis, while the corresponding y-axis illustrates the frequency or count of occurrences.

- This visualization provides a clear and concise overview of the prevalence of different intrusion categories, allowing for quick identification of major threats.

- Bar graphs are particularly effective in conveying the relative frequencies of attacks, aiding cybersecurity professionals in prioritizing their focus on the most prominent security concerns.

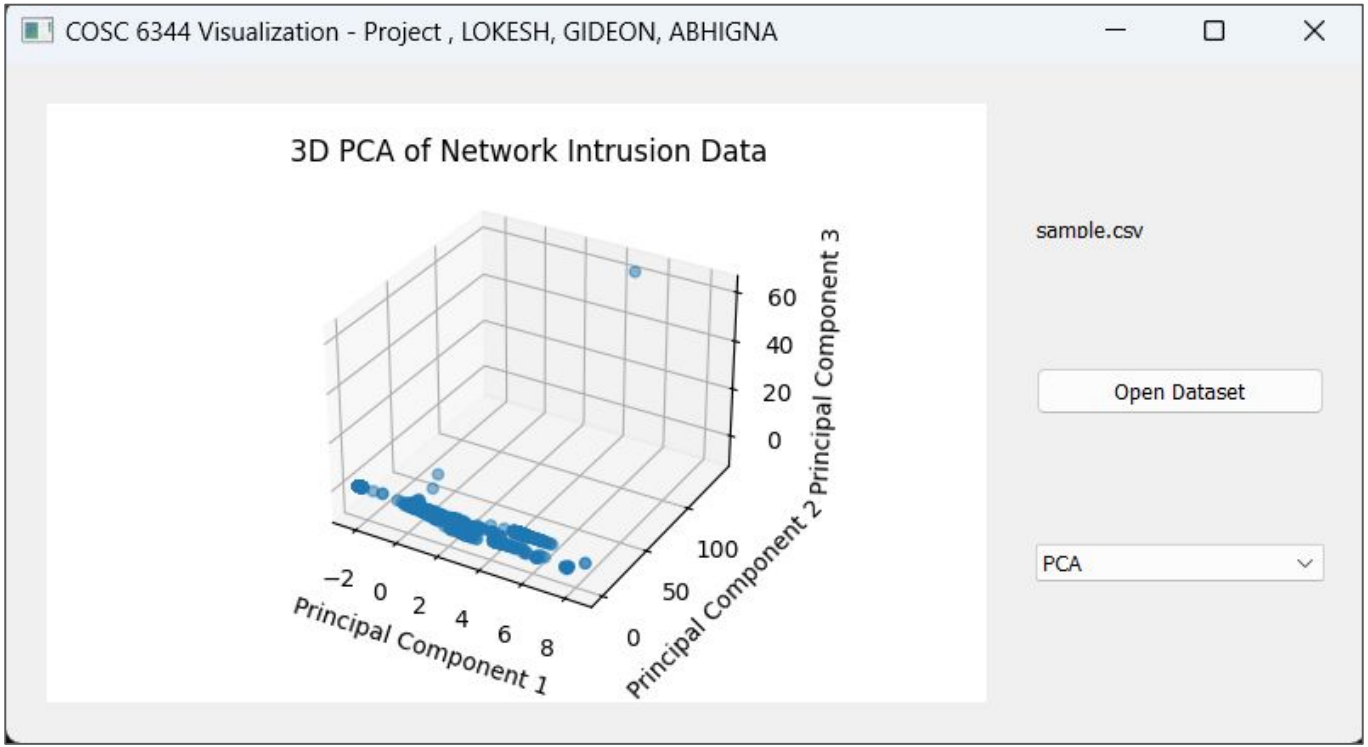
PIE CHARTS

- Each slice of the pie corresponds to a specific attack category, with the size of each slice proportional to the percentage of instances it represents.

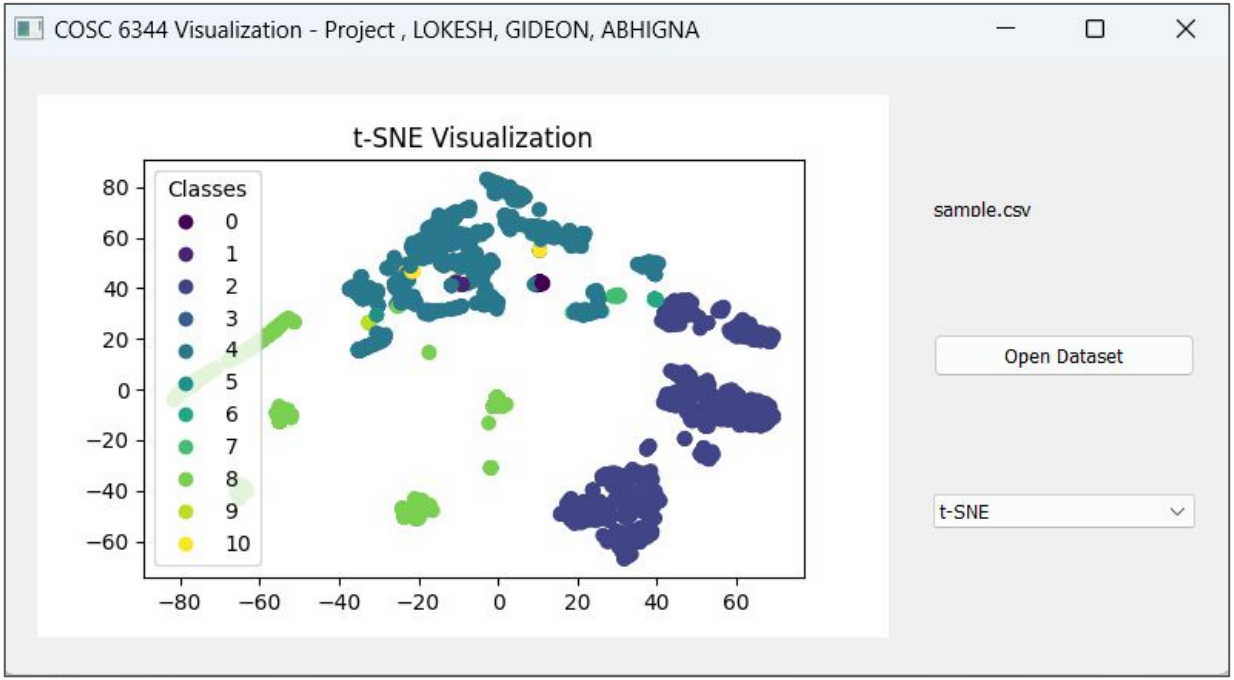
- Pie charts provide an intuitive and visually appealing representation of the dataset's composition, enabling stakeholders to grasp the relative importance of each attack type at a glance.

- This visualization technique adds a layer of accessibility to the analysis, supporting a broader audience in understanding the distribution of network intrusions.

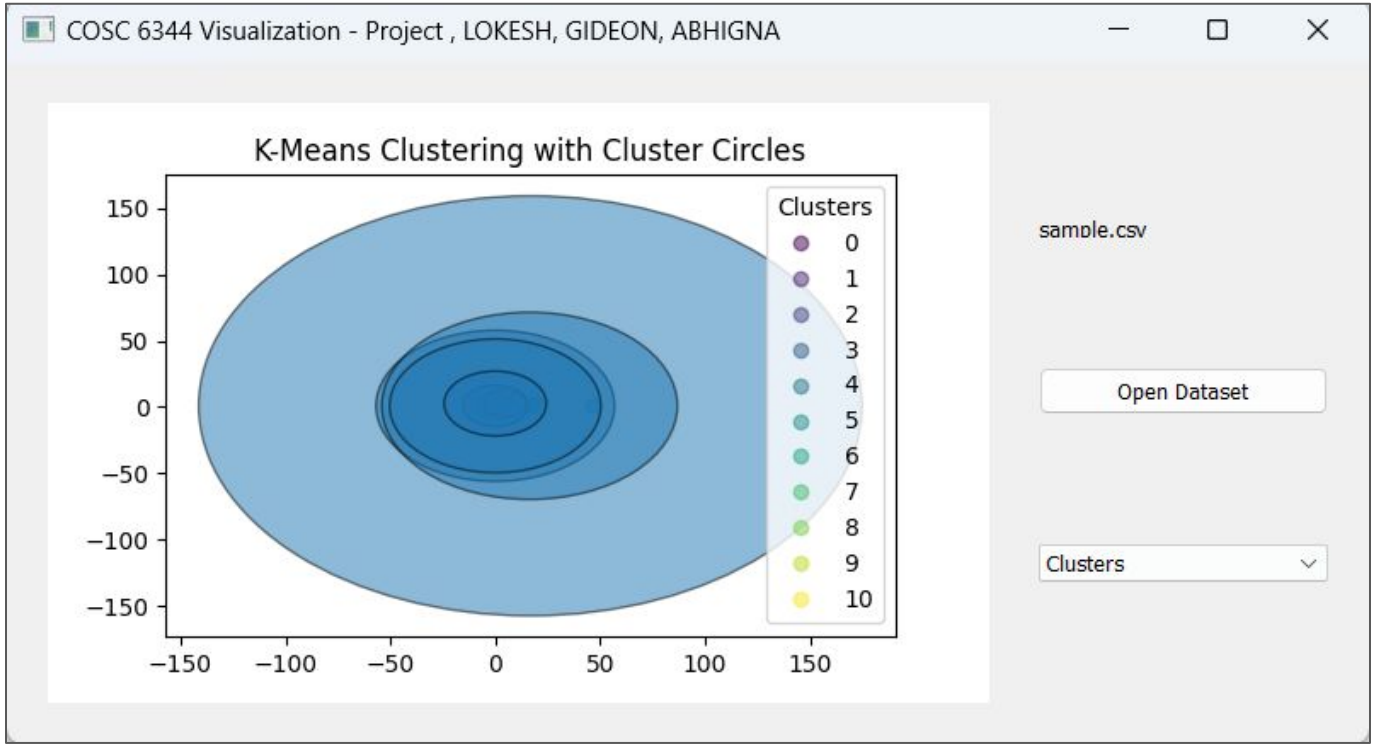
RESULTS : Principal Component Analysis (PCA)



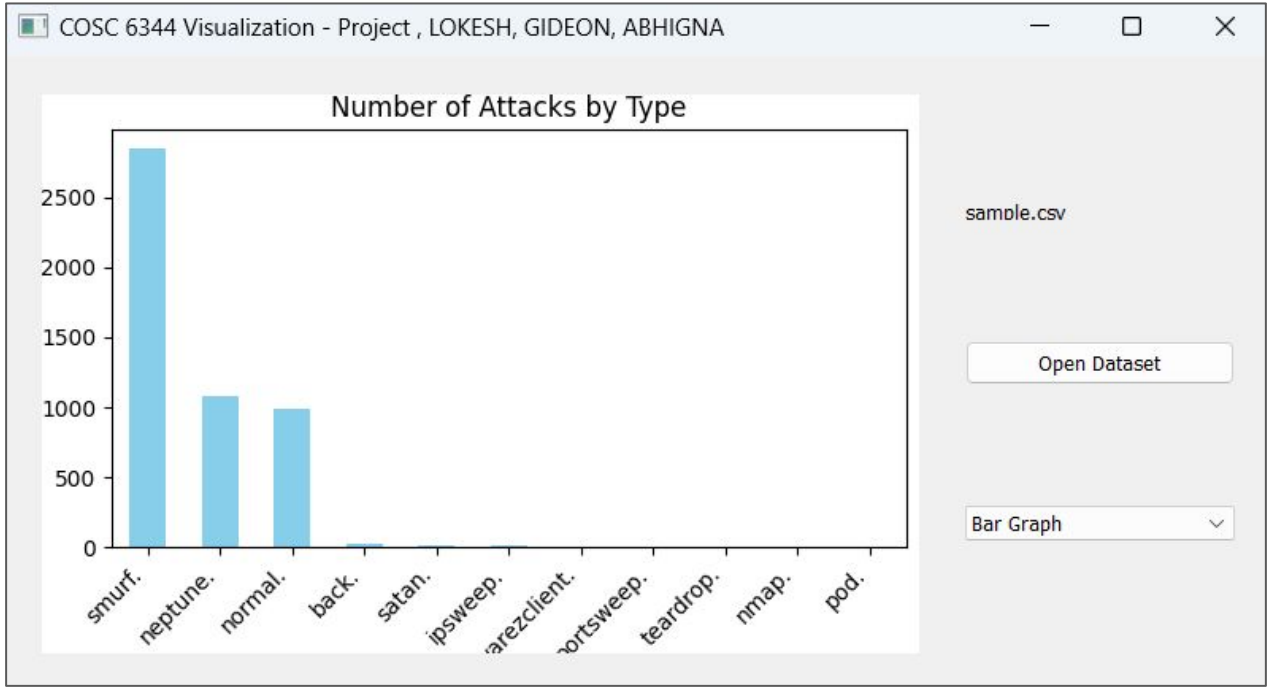
RESULTS : t-distributed Stochastic Neighbor Embedding (t-SNE)



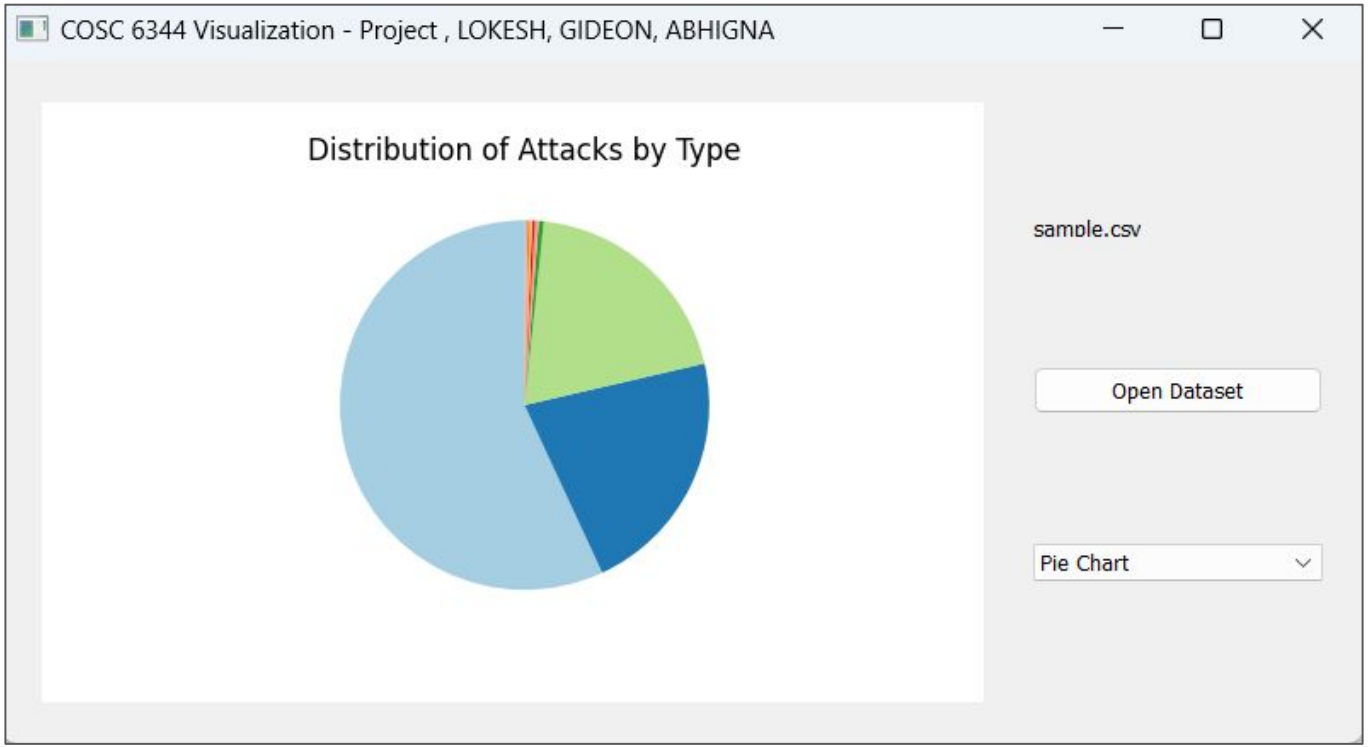
RESULTS : Clusters of Attack Types - k-means



RESULTS - Bar Graph - Attack Type vs Number of Attacks



RESULTS - Pie Chart - Attack Types



CONCLUSION

Attacks types and the complexity of the dimensions are visualized using PCA, t-sne, clustering, bar graphs, and pie charts.

Future Work:

Adding functionality for more types of visualizations.

Improving the dataset by doing better preprocessing to discover hidden features.

Thank You