

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/329955939>

# A 3D Approach for the Visualization of Network Intrusion Detection Data

Conference Paper · October 2018

DOI: 10.1109/CW.2018.00064

---

CITATIONS

8

READS

174

3 authors:



Wei Zong

University of Wollongong

16 PUBLICATIONS 162 CITATIONS

[SEE PROFILE](#)



Yang-Wai Chow

University of Wollongong

95 PUBLICATIONS 943 CITATIONS

[SEE PROFILE](#)



Willy Susilo

University of Wollongong

919 PUBLICATIONS 23,407 CITATIONS

[SEE PROFILE](#)

# A 3D Approach for the Visualization of Network Intrusion Detection Data

Wei Zong, Yang-Wai Chow, Willy Susilo

*Institute of Cybersecurity and Cryptology*

*School of Computing and Information Technology*

*University of Wollongong, Australia*

*Email:* wz630@uowmail.edu.au, caseyc@uow.edu.au, wsusilo@uow.edu.au

**Abstract**—With the increasing threat of cyber attacks, machine learning techniques have been researched extensively in the area of network intrusion detection. Such techniques can potentially provide a means for the real-time automated detection of attacks and abnormal traffic patterns. However, misclassification is a common problem in machine learning techniques for intrusion detection, and a lack of insight into why such misclassification occurs impedes the improvement of machine learning models. This paper presents an approach to visualizing network intrusion detection data in 3D. The purpose of this is to facilitate the understanding of network intrusion detection datasets using a visual representation to reflect the geometric relationship between various categories of network traffic. This can potentially provide useful insight to aid the design of machine learning techniques. This paper demonstrates the usefulness of the proposed 3D visualization approach by presenting results of experiments on commonly used network intrusion detection datasets.

**Keywords**-network intrusion detection; machine learning; principal component analysis; visualization;

## I. INTRODUCTION

It is well known that Machine Learning (ML) techniques have long been used for analyzing and extracting useful information from data [1]. With the increasing threat of cyber attacks nowadays, cyber security experts have undertaken wide ranging studies on techniques for combating such security threats. Network Intrusion Detection Systems (NIDS) are potential automated solutions for protecting online environments. As such, ML techniques have been researched extensively in the area of NIDS as a promising solution for automating the real-time detection of attacks or abnormal traffic patterns in a network [2], [3]. For this reason, it is vital to understand the significance and reasons behind intrusion detection results that are produced by ML models [3].

Misclassification is a common problem in ML. A lack of insight into why such misclassification occurs impedes the development and improvement of ML models. While many studies on using ML in NIDS have presented improved results when compared with other techniques, there is not much emphasis on understanding the reasons for the improved results or lower misclassification rates [4], [5], [6]. These studies tend to present numerical results in tables or graphs to compare performances between different ML

techniques.

However, the underlying reasons for poor performance in the detection of certain attack categories are not usually explained and cannot be perceived clearly or intuitively. Without a comprehensible approach to analyzing the reasons for poor performance, the improvement of ML models usually relies on a trial-and-error process due to the complex nature of ML mechanisms [7]. More specifically, the design of ML models heavily depends on the characteristics of the training and testing datasets. Hence, an intuitive and explainable analysis on such datasets could potentially provide insight on ways of improving current ML techniques.

Public network intrusion detection datasets, such as the NSL\_KDD [8] and UNSW-NB15 [9] datasets, are commonly used as benchmark datasets for validating ML techniques for NIDS. Information visualization can play an important role in analyzing datasets. This is because dataset visualization can facilitate mental perception and provide insight into complex data structures. Furthermore, compared to numeric data presentation, visual representation is more inspiring and intuitive [10].

Various forms of visualization techniques have already been applied successfully in the cyber security community, ranging from 2-dimensional (2D) to 3-dimensional (3D) applications [11], [12]. Existing research in cyber security visualization covers a variety of domains, including for monitoring network traffic characteristics [13], [14] and for visualizing complex attack patterns [15], [16], [17]. While studies have performed analysis on NIDS datasets [8], [9], there is currently not much research on visualization systems for providing visual representation of NIDS data, which can provide insight and improve the design of a variety of NIDS ML models [7].

This paper proposes a 3D visualization technique for analyzing NIDS datasets. The purpose of the proposed approach is to provide a visual representation of data from NIDS datasets, in a way which reflects the geometric relationship between diverse network traffic data records, and to create a way of examining the likely causes of ML misclassification from a visual perspective. In addition, this paper demonstrates results of a system that was developed based on the proposed visualization approach.

## II. BACKGROUND

This section presents a background to topics that are related to the work in this research.

### A. Information Gain

Feature selection is an important step in building ML models. This is because only certain important features in the data may be essential to the classification process. The traditional approach is where security experts rank the importance of features manually. However, it would be ideal if an automated approach to selecting important features could be used.

Information gain is an approach to this, where features with low information gain can be eliminated. This is because they are considered to be unimportant as they have relatively small relevance on classification. Information gain can only be calculated using discrete variables, and it is equal to subtracting the sum of entropy for each subset of records, weighted by their probability of occurring, from the entropy of the target feature of the original dataset. A method of calculating information gain is provided as follows [18]:

Let  $X$  and  $Y$  be discrete variables representing sample attributes  $(x_1, x_2, \dots, x_m)$  and class attributes  $(y_1, y_2, \dots, y_n)$ , respectively. Then, the information gain,  $IGain$ , of a given attribute  $X$  regarding a class attribute  $Y$  is calculated as:

$$IGain(Y, X) = Entropy(Y) - Entropy(Y|X) \quad (1)$$

where

- $Entropy(Y) = -\sum_{i=1}^n P(Y = y_i) \log_2 P(Y = y_i)$ , where  $P(Y = y_i)$  is the probability that  $y_i$  occurs, and
- $Entropy(Y|X) = -\sum_{i=1}^m P(X = x_j) Entropy(Y|X = x_j)$ .

### B. Network Intrusion Detection Datasets

Network intrusion detection datasets are essential for the development and evaluation of the effectiveness of various techniques used in NIDS. Benchmark datasets that are commonly used by the research community include the KDD98, KDD\_CUP99 and NSL\_KDD datasets. These datasets contain a number of different categories of network traffic including normal traffic and attacks, such as Denial of Service (DoS), probe attacks, Remote to Local (R2L) and User to Root (U2R) attacks. However, it has been contended that these network intrusion detection datasets were generated more than a decade ago, and several studies have highlighted flaws in these datasets [8], [19].

In light of this, the UNSW-NB15 dataset was proposed as a contemporary dataset that was created as a hybrid of modern normal and contemporary synthesized attack activities of network traffic [9]. This dataset breaks down the attack categories into other attacks like worms, shellcode, exploits, fuzzers, etc. While experiments conducted in this study focused on using the NSL\_KDD and the UNSW-NB15

datasets, the proposed approach is a generic visualization method that can also handle other datasets.

## III. RELATED WORK

The PCA technique is an important tool that has been used in a variety of work, including for research on analyzing network traffic for NIDS. Moustafa et al. [6] adopted the PCA technique to reduce high dimensionality of network connections. Hoz Correa et al. [20] used PCA for feature selection and noise removal in NIDS, while Lakhina et al. [21] proposed to use PCA as a statistical tool in network anomaly detection. In their approach, network traffic data was divided into normal and abnormal subspaces, before statistical analysis was performed to detect anomalies. In other work, Camacho et al. [22] introduced the main steps of the Multivariate Statistical Process Control approach for NIDS based on the PCA technique. They also suggested that before applying PCA, the importance of each feature on the classification process should be considered.

As ML is a promising solution for providing an automated mechanism for network intrusion detection, there is a much interest and a variety of research on this topic. For example, Lin et al. [5] proposed a novel feature representation method, which considers the geometric properties of datasets and uses the k-Nearest Neighbor (kNN) classifier to detect network attacks. Although their approach performs well in detecting normal traffic, DoS and probe attacks, the detection accuracy of their approach is not satisfactory for other categories like U2R and R2L attacks.

Wang et al. [4] proposed an approach that first used a fuzzy clustering technique to divide the training set into several subsets. Then different Artificial Neural Networks were trained on these subsets. Finally, a fuzzy aggregation module was used to combine the detection results. Their experimental results demonstrated high performance on intrusion detection using this multi-stage approach. Moustafa et al. [6] propose a novel technique, called Geometric Area Analysis based on Trapezoidal Area Estimation for NIDS. While this approach is effective in detecting intrusions for both the NSL\_KDD and UNSW-NB15 datasets, the reasons for misclassification are not really described in detail.

Many studies usually do not describe the reasons for ML misclassification. This is in part due to the incomprehensible nature of ML algorithms, where many users regard ML as a black box. Nevertheless, instead of using trial-and-error process, visualization has been proposed as a critical tool for improving ML models. This trend has attracted many researchers, which has resulted in various studies on the topic [7]. As an example, Rauber et al. [23] presented work to visualize relationships between learned representations of observations, and relationships between artificial neurons to give network designers highly valuable insight into how their systems operate. In addition, Liu et al. [7] proposed a visual

analytics system which facilitates understanding, diagnosing and refining deep convolution neural networks.

There have also been a number of visualization approaches proposed in the NIDS domain. Yelizarov et al. [15] presented a visual technique, which can be used to display the overall status of the network and to show complex attack patterns. A cyber security visualization system for contributing to situation awareness by helping users to understand the cyber security status and events, has also been proposed [16]. McKenna et al. [17] designed a cyber security dashboard to help network analysts in identifying and summarizing patterns within network data.

In other work, Onut et al. [24] projected network data in 3D space and try to distinguish different attacks in a visual manner. Examples of several attack scenarios were described and discussed in their work. A NIDS based on 2D visual presentation of network data was proposed by Corchado et al. [25]. In their work, network traffic was visualized based on 5 variables of each packet, namely, source port, destination port, size, timestamp and protocol. In their system, anomalies in network traffic showed up as different patterns compared to normal connections.

Despite the many efforts in the development of visualization methods in cyber security, visualization techniques aimed at improving ML models for NIDS are scarce. The work in this paper attempts to address this.

#### IV. PROPOSED APPROACH

The proposed 3D approach for the visualization of network intrusion detection data for machine learning is presented in this section. Figure 1 provides an overall depiction of the stages in the proposed approach.

The first stage involves the extraction of network traffic records from the dataset. In view of the fact that in each network intrusion detection dataset, minor categories like worm attacks (in the UNSW-NB15 dataset), and U2R and R2L attacks (in the NSL\_KDD dataset), only occupy a small portion of the dataset, all the data for the minor categories are extracted from both the training and testing sets. The data from the remaining major categories are randomly extracted until a certain predefined amount, e.g., 30% of the data. While the full dataset can be used, the random sampling of data from the major categories does not diminish the quality of visual information, but it is useful to reduce the visual clutter and the amount of required computation on the visualization system.

The next stage of the process is to apply one-hot transformation. PCA is not suitable for categorical data, as such one-hot transform is applied to the features to transform them into binary features. The transformation is done on the training and testing sets individually. As a result, all the possible features will be created because some features may appear in the training set but not in the testing set. Features that only appear in the testing set will be ignored. This is

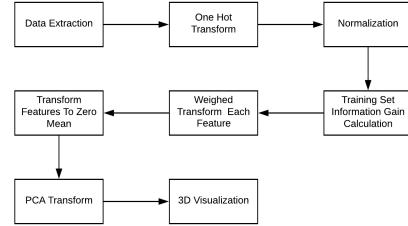


Figure 1. Stages in the proposed approach.

followed by a normalization process to normalize the range of the features. The purpose of performing normalization is because the range of values for the features can differ significantly. For example, some values may range between zero to less than a hundred, while others may range from 0 to several millions. To correctly utilize PCA, linear normalization is applied to each feature in both the training and testing sets to normalize the values to the minimum and maximum range of the training set.

As previously mentioned, Camacho et al. [22] suggested that before applying PCA, the importance of each feature in relation to classification should be considered. Thus, the information gain of each feature is calculated to ascertain the importance of each feature of the training set. Since information gain requires discrete features, equal frequency binning discretization is first performed. Then, the information gain values are normalized between 0 and 1, and weighted transform is applied by multiplying each feature with the corresponding information gain value. As a result, the variances of important features either remain or are decreased slightly, whereas the variances of unimportant features are significantly reduced.

To correctly apply PCA transformation, the mean of each feature must be zero. Hence, all features are transformed to zero mean in the training set, and the testing set is transformed based on mean values of the corresponding features from the training set. The PCA model is trained on the training set and then used to transform both the training and testing sets. The resulting variance of the first 16 principal components for the NSL\_KDD and UNSW-NB15 datasets is shown in Figure 2(a) and Figure 2(b), respectively. It should be noted that the first 13 components are the most significant, representing 96.3% of the variances for the UNSW-NB15 dataset and 98.7% for the NSL\_KDD dataset.

Finally, to visualize these 13 components in 3D space, 13 non-parallel vectors are defined. The  $(x, y, z)$  components of the 13 non-parallel vectors are provided in Table I. The resulting 3D coordinate of a network traffic record is computed as a weighted sum of its 13 PCA components,  $c_1, c_2, \dots, c_{13}$ , with the normalized unit normal vectors of the 13 non-parallel vectors,  $v_1, v_2, \dots, v_{13}$ . The result is multiplied by a constant scaling factor,  $s$ , which is used

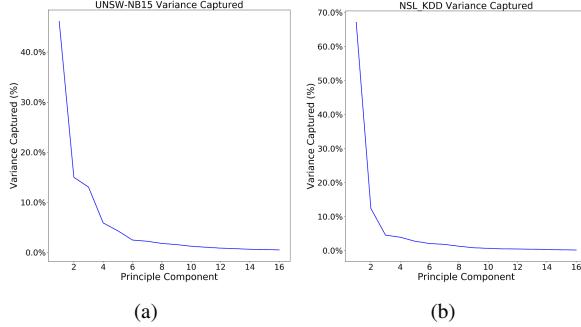


Figure 2. Variance of the first 16 principal components; (a) for the UNSW-NB15 dataset; (b) for the NSL\_KDD dataset.

Table I  
( $x, y, z$ ) COMPONENTS OF THE 13 NON-PARALLEL VECTORS.

	( $x, y, z$ ) components
$v_1$	(1, 0, 0)
$v_2$	(0, 1, 0)
$v_3$	(0, 0, 1)
$v_4$	(1, 1, 0)
$v_5$	(1, 0, 1)
$v_6$	(0, 1, 1)
$v_7$	(1, -1, 0)
$v_8$	(1, 0, -1)
$v_9$	(0, 1, 1)
$v_{10}$	(1, 1, 1)
$v_{11}$	(1, 1, -1)
$v_{12}$	(1, -1, 1)
$v_{13}$	(-1, 1, 1)

to control the degree of separation between positions. The formula for the weighted sum is as follows:

$$position = \sum_{i=1}^{13} c_i \hat{v}_i s \quad (2)$$

Spheres with different colors are then rendered at these positions to represent different categories of network traffic.

## V. RESULTS AND DISCUSSION

This section demonstrates results obtained from the system that was developed based on the proposed approach. The UNSW-NB15 and NSL\_KDD datasets were used to illustrate the results, and some representative observations from these datasets are presented here.

### A. The UNSW-NB15 Dataset

A 3D visual representation of the UNSW-NB15 dataset as displayed by the system is shown in Figure 3. Figure 3(a) shows a visual depiction of a portion of the training set, where the different network traffic categories have been color coded to visually distinguish them from one another. It can be seen visually that traffic from the same category are typically clustered together. The visual representation of part of the testing set is shown in Figure 3(b). When comparing

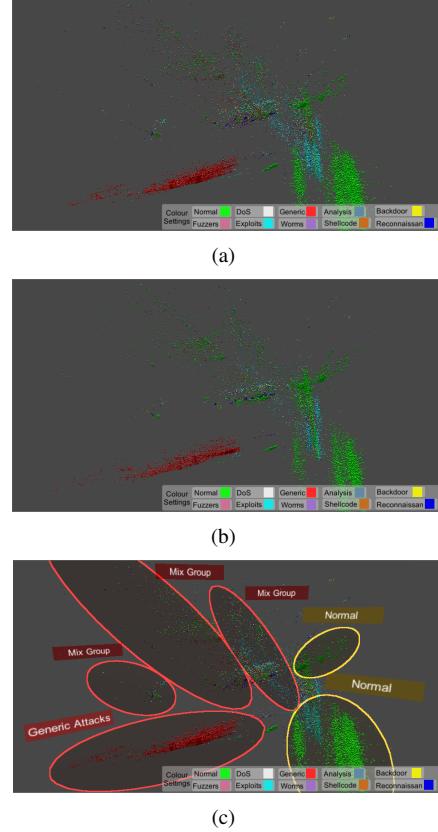


Figure 3. 3D visual representation of the UNSW-NB15 dataset; (a) data from the training set; (b) data from the testing set; (c) empirical grouping of data from the training set.

the visual representation of the training set with the testing set, it can clearly be seen that the characteristics of both sets are similar. This means that due to the similar patterns, when training a ML model using the training set it is highly likely that the model will be able to detect attacks in the testing set.

Figure 3(c) provides an empirical grouping of data from the training set. The groups circled in yellow depict sections that mainly only contain normal network traffic, whereas other groupings are circled in red. It can be seen from the groupings that the traffic for generic attacks are well clustered together. However, the traffic in the mixed groups is not so well defined as they contain both attacks and normal network traffic.

Figure 4 provides examples of certain groups for closer examination. Examples of normal traffic clusters from the training and testing data are shown in Figure 4(a) and Figure 4(b), respectively. It can be seen from the figures that the characteristics of the training and testing data are very similar and that there are three clusters of normal traffic within the group. The implication of this for ML is that traffic in groups that contain nearly homogeneous records are easier to correctly identify and typically result in high

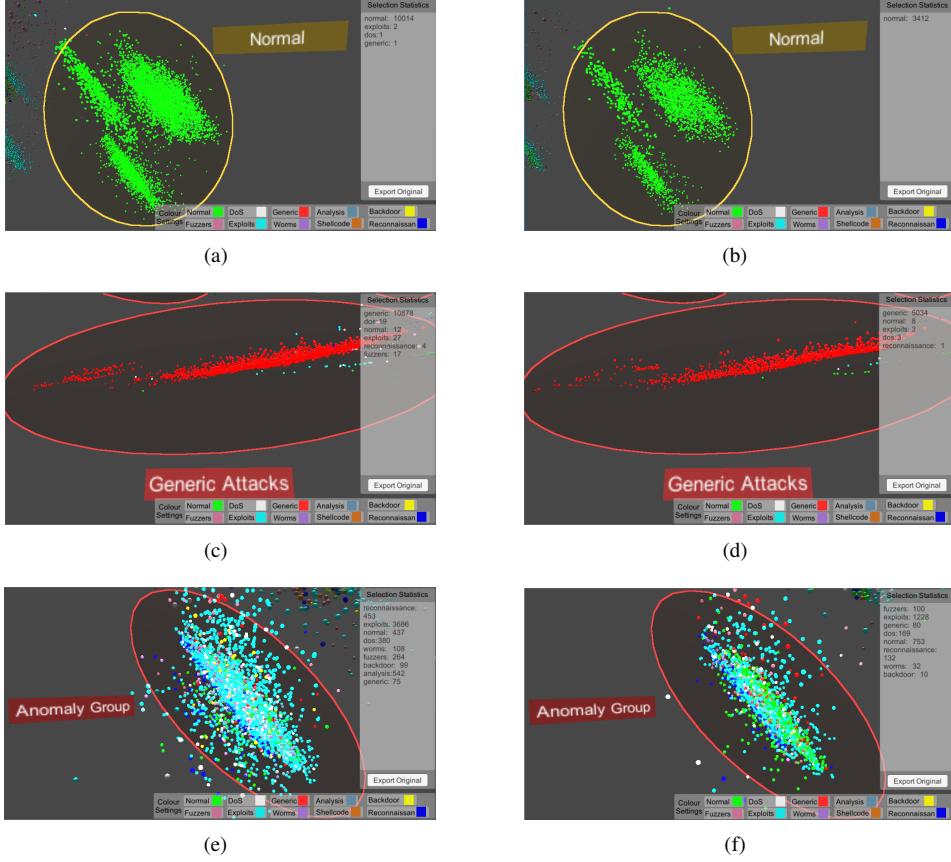


Figure 4. Example groups from the UNSW-NB15 dataset; (a)-(b) a group containing normal traffic from the training and testing data, respectively; (c)-(d) a group containing mostly generic attacks from the training and testing data, respectively; (e)-(f) a group containing mixed traffic from the training and testing data.

detection performance. This is because their characteristics have similar patterns with little variation between them, which results in them being clustered together. Figure 4(c) and Figure 4(d) show examples of a group from the training and testing data, respectively, that contain mostly generic attacks. Note that the statistics of the traffic in the group is shown in the panel on the right. Since the group contains mainly homogeneous records, it can be concluded that in general the detection performance of most ML models on these generic attacks will be high. The results from several studies appear to support this conclusion [6], [26].

However, in other sections of the dataset, the traffic is mixed together and there is no clear visual distinction between the different traffic categories within these groups. An example of such a group is shown in Figure 4(e) and Figure 4(f) for the training and testing data, respectively. Such sections of the dataset are where ML techniques tend to face difficulties when it comes to the task of correctly identifying the individual categories of the traffic. In addition, the figures provide a visual comparison of the data between the training and testing set. It can be seen that the training set contains a number of analysis attacks, but there are none in the

testing set. There is a low number of worm attacks, which is disproportional when comparing the numbers in the training set with the testing set. Furthermore, as there is a significant number of exploit attacks within the group, it is challenging for ML models to avoid misclassifying other traffic in the group as exploits.

Therefore, to address the problem of highly mixed groups, a feature representation approach like the CANN method [5], can potentially be used to transform these records before the ML model is trained and also before using the trained model for detection. In view of the fact that the training and testing sets exhibit similar patterns, as can be seen from Figure 3(a) and Figure 3(b), it can be anticipated that if a feature representation approach successfully transforms the training data into groups that contain mainly homogeneous records, the same approach should be able to transform the testing records into corresponding homogeneous groups.

#### B. The NSL\_KDD Dataset

Figure 5 shows the overall 3D visual representation of the NSL\_KDD dataset as displayed by the proposed system. Figure 5(a) depicts a portion of the data from the training

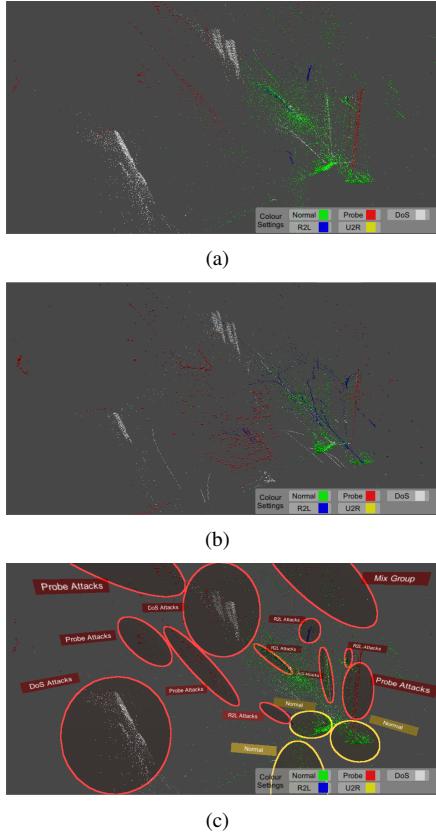


Figure 5. 3D visual representation of the NSL\_KDD dataset; (a) data from the training set; (b) data from the testing set; (c) empirical grouping of data from the training set.

set, while Figure 5(b) displays part of the data from the testing set. It can be seen that while the overall visual distributions share relatively similar characteristics, the training and testing sets in the NSL\_KDD dataset have more differences between them as compared with the UNSW-NB15 dataset (i.e. the characteristics of traffic between the training and testing sets in the UNSW-NB15 dataset exhibited higher similarities). Figure 5(c) in turn shows the empirical grouping of data from the training set. Like the UNSW-NB15 dataset, it also shows that the data can be grouped into several clusters.

From the visual representation, one can observe that the data consists of groups that contain mainly homogeneous records as well as groups that contain diverse traffic. Figure 6(a) and Figure 6(b) show examples of DoS attack traffic from the training and testing data, respectively. It can clearly be seen that the display of traffic within the group is isolated away from the rest of the traffic, and contains mostly homogeneous DoS attack records. Similarly, another example of a group containing highly distinguishable homogeneous traffic is one containing probe attacks, as shown in Figure 6(c) and Figure 6(d) from the training and testing data, respectively. Due to the isolation and clustering of such homogeneous

data, ML techniques are expected to perform well when identifying such traffic.

However, unlike the previous two examples, the main difficulty faced by ML techniques in the NSL\_KDD comes from previously unknown attacks in the testing set [27]. Even though the categories of attacks are the same in the training and testing sets, the characteristics of these traffic significantly differ. A visual representation of this comparing the training and testing data is shown in Figure 6(e) and Figure 6(f), respectively. It can be seen that in the training set in Figure 6(e) that there are no attacks within this group. On the other hand, there are many R2L and DoS attacks in the corresponding group from the testing set, as shown in Figure 6(f). Despite the fact that ML models should be designed based on the training set, ML techniques that do not consider such abnormal characteristics may perform unsatisfactorily when it comes to detecting such attacks.

It should be noted that the proposed 3D visualization system is an interactive system that is displayed in real-time. The user can freely navigate the virtual camera in 3D space to examine different areas of the display, to obtain statistical information and also to extract selected data from the visual representation for closer inspection. While the current system deals with data from a network intrusion detection dataset, the end goal is for the system to be able to handle live network traffic. This will allow cyber security experts to examine and recognize patterns of incoming network traffic in real-time, and will be the focus of future work.

The experiments conducted in this study show different geometric characteristics in the visual representation of the UNSW-NB15 and NSL\_KDD datasets. It was highlighted that the main challenge presented to ML techniques in the UNSW-NB15 dataset is due to sections that contain groups with highly heterogeneous data, whereas the main challenge in the NSL\_KDD dataset is due to the abnormal characteristics of unknown attacks. This demonstrates the usefulness of the proposed 3D visualization approach in identifying patterns and clusters in the data. In order to address these issues and improve detection performance as well as to reduce misclassification rates, it is important for NIDS ML techniques to consider these geometric characteristics in their design.

## VI. CONCLUSION

This paper presents an approach to visualizing network intrusion detection data in 3D. The aim of the proposed approach is to facilitate the understanding of NIDS datasets using a visual representation to reflect the geometric relationship between various categories of network traffic. This can provide useful insight for the design of machine learning techniques, such as to understand the reasons for high misclassification rates in certain machine learning models.

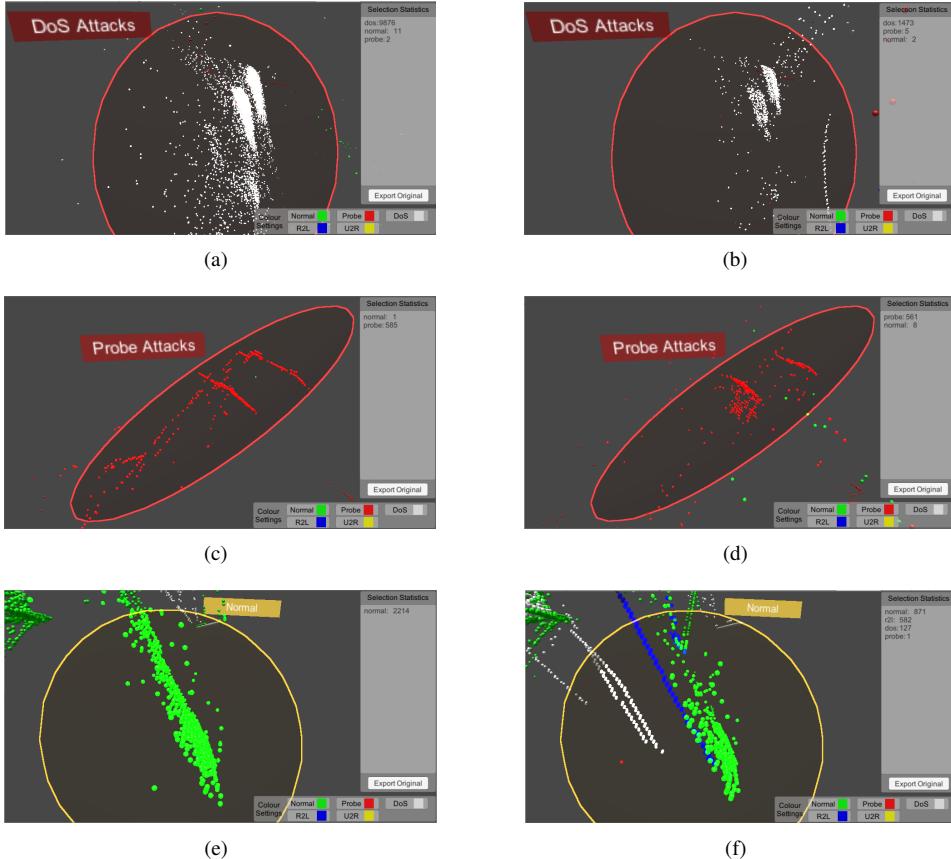


Figure 6. Examples of groups from the NSL\_KDD dataset; (a)-(b) a group containing DoS attacks from the training and testing data, respectively; (c)-(d) a group containing mostly probe attacks from the training and testing data, respectively; (e)-(f) a group containing mainly normal traffic from the training and testing data, respectively.

A system was developed based on the proposed visualization approach, and results of experiments on commonly used NIDS datasets were presented. This demonstrates the usefulness of the proposed 3D visualization approach in identifying patterns and clusters in the data. Future work will focus on the visualization of live network traffic to allow users to examine incoming network traffic in a real-time interactive manner.

#### REFERENCES

- [1] S. Suthaharan, "Big data classification: problems and challenges in network intrusion prediction with machine learning," *SIGMETRICS Performance Evaluation Review*, vol. 41, no. 4, pp. 70–73, 2014. [Online]. Available: <http://doi.acm.org/10.1145/2627534.2627557>
- [2] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016. [Online]. Available: <https://doi.org/10.1109/COMST.2015.2494502>
- [3] R. Sommer and V. Paxson, "Outside the closed world: On using machine learning for network intrusion detection," in *31st IEEE Symposium on Security and Privacy, S&P 2010*, 16-19 May 2010, Berkeley/Oakland, California, USA. IEEE Computer Society, 2010, pp. 305–316. [Online]. Available: <https://doi.org/10.1109/SP.2010.25>
- [4] G. Wang, J. Hao, J. Ma, and L. Huang, "A new approach to intrusion detection using artificial neural networks and fuzzy clustering," *Expert Syst. Appl.*, vol. 37, no. 9, pp. 6225–6232, 2010. [Online]. Available: <https://doi.org/10.1016/j.eswa.2010.02.102>
- [5] W. Lin, S. Ke, and C. Tsai, "CANN: an intrusion detection system based on combining cluster centers and nearest neighbors," *Knowl.-Based Syst.*, vol. 78, pp. 13–21, 2015. [Online]. Available: <https://doi.org/10.1016/j.knosys.2015.01.009>
- [6] N. Moustafa, J. Slay, and G. Creech, "Novel geometric area analysis technique for anomaly detection using trapezoidal area estimation on large-scale networks," *IEEE Transactions on Big Data*, p. 1. [Online]. Available: [doi.ieeecomputersociety.org/10.1109/TBDA.2017.2715166](https://doi.ieeecomputersociety.org/10.1109/TBDA.2017.2715166)
- [7] S. Liu, X. Wang, M. Liu, and J. Zhu, "Towards better analysis of machine learning models: A visual analytics perspective," *Visual Informatics*, vol. 1, no. 1, pp. 48–56, 2017. [Online]. Available: <https://doi.org/10.1016/j.visinf.2017.01.006>

- [8] M. Tavallaei, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009, Ottawa, Canada, July 8-10, 2009*. IEEE, 2009, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/CISDA.2009.5356528>
- [9] N. Moustafa and J. Slay, "The evaluation of network anomaly detection systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set," *Information Security Journal: A Global Perspective*, vol. 25, no. 1-3, pp. 18–31, 2016. [Online]. Available: <https://doi.org/10.1080/19393555.2015.1125974>
- [10] S. Liu, W. Cui, Y. Wu, and M. Liu, "A survey on information visualization: recent advances and challenges," *The Visual Computer*, vol. 30, no. 12, pp. 1373–1393, Dec 2014. [Online]. Available: <https://doi.org/10.1007/s00371-013-0892-3>
- [11] H. Shiravi, A. Shiravi, and A. A. Ghorbani, "A survey of visualization systems for network security," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 8, pp. 1313–1329, 2012. [Online]. Available: <https://doi.org/10.1109/TVCG.2011.144>
- [12] D. Staheli, T. Yu, R. J. Crouser, S. Damodaran, K. Nam, B. D. O'Gwynn, S. McKenna, and L. Harrison, "Visualization evaluation for cyber security: trends and future directions," in *Proceedings of the Eleventh Workshop on Visualization for Cyber Security, Paris, France, November 10, 2014*, K. Whitley, S. Engle, L. Harrison, F. Fischer, and N. Prigent, Eds. ACM, 2014, pp. 49–56. [Online]. Available: <http://doi.acm.org/10.1145/2671491.2671492>
- [13] R. Ball, G. A. Fink, and C. North, "Home-centric visualization of network traffic for security administration," in *Workshop on Visualization and Data Mining for Computer Security (VizSEC/DMSEC 2004), 29 October 2004, Washington DC, USA*, C. E. Brodley, P. Chan, R. Lippmann, and W. Yurcik, Eds. ACM, 2004, pp. 55–64. [Online]. Available: <http://doi.acm.org/10.1145/1029208.1029217>
- [14] J. R. Goodall, W. G. Lutters, P. Rheingans, and A. Komlodi, "Preserving the big picture: Visual network traffic analysis with TN," in *IEEE Workshop on Visualization for Computer Security (VizSEC 2005), 26 October 2005, Minneapolis, MN, USA*, K. Ma, S. C. North, and W. Yurcik, Eds. IEEE Computer Society, 2005, p. 6. [Online]. Available: <https://doi.org/10.1109/VIZSEC.2005.17>
- [15] A. Yelizarov and D. Gamayunov, "Visualization of complex attacks and state of attacked network," in *6th International Workshop on Visualization for Cyber Security 2009, VizSec 2009, Atlantic City, New Jersey, USA, October 11, 2009*, D. A. Frincke, C. Gates, J. R. Goodall, and R. F. Erbacher, Eds. IEEE Computer Society, 2009, pp. 1–9. [Online]. Available: <https://doi.org/10.1109/VIZSEC.2009.5375527>
- [16] M. Angelini, N. Prigent, and G. Santucci, "PERCIVAL: proactive and reactive attack and response assessment for cyber incidents using visual analytics," in *2015 IEEE Symposium on Visualization for Cyber Security, VizSec 2015, Chicago, IL, USA, October 25, 2015*, L. Harrison, N. Prigent, S. Engle, and D. M. Best, Eds. IEEE Computer Society, 2015, pp. 1–8. [Online]. Available: <https://doi.org/10.1109/VIZSEC.2015.7312764>
- [17] S. McKenna, D. Staheli, C. Fulcher, and M. D. Meyer, "Bubblenet: A cyber security dashboard for visualizing patterns," *Comput. Graph. Forum*, vol. 35, no. 3, pp. 281–290, 2016. [Online]. Available: <https://doi.org/10.1111/cgf.12904>
- [18] P. Sangkatsanee, N. Wattanapongsakorn, and C. Charnsripinyo, "Practical real-time intrusion detection using machine learning approaches," *Computer Communications*, vol. 34, no. 18, pp. 2227–2235, 2011. [Online]. Available: <https://doi.org/10.1016/j.comcom.2011.07.001>
- [19] J. McHugh, "Testing intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by lincoln laboratory," *ACM Trans. Inf. Syst. Secur.*, vol. 3, no. 4, pp. 262–294, 2000. [Online]. Available: <http://doi.acm.org/10.1145/382912.382923>
- [20] E. de la Hoz Correa, E. de la Hoz Franco, A. Ortiz, J. Ortega, and B. Prieto, "PCA filtering and probabilistic SOM for network intrusion detection," *Neurocomputing*, vol. 164, pp. 71–81, 2015. [Online]. Available: <https://doi.org/10.1016/j.neucom.2014.09.083>
- [21] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," in *Proceedings of the ACM SIGCOMM 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication, August 30 - September 3, 2004, Portland, Oregon, USA*, R. Yavatkar, E. W. Zegura, and J. Rexford, Eds. ACM, 2004, pp. 219–230. [Online]. Available: <http://doi.acm.org/10.1145/1015467.1015492>
- [22] J. Camacho, A. Pérez-Villegas, P. García-Teodoro, and G. Maciá-Fernández, "Pca-based multivariate statistical network monitoring for anomaly detection," *Computers & Security*, vol. 59, pp. 118–137, 2016. [Online]. Available: <https://doi.org/10.1016/j.cose.2016.02.008>
- [23] P. E. Rauber, S. G. Fadel, A. X. Falcão, and A. C. Telea, "Visualizing the hidden activity of artificial neural networks," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 101–110, 2017. [Online]. Available: <https://doi.org/10.1109/TVCG.2016.2598838>
- [24] I. Onut and A. A. Ghorbani, "Svision: A novel visual network-anomaly identification technique," *Computers & Security*, vol. 26, no. 3, pp. 201–212, 2007. [Online]. Available: <https://doi.org/10.1016/j.cose.2006.10.001>
- [25] E. Corchado and Á. Herrero, "Neural visualization of network traffic data for intrusion detection," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2042–2056, 2011. [Online]. Available: <https://doi.org/10.1016/j.asoc.2010.07.002>
- [26] T. Janarthanan and S. Zargari, "Feature selection in unsw-nb15 and kddcup'99 datasets," in *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, June 2017, pp. 1881–1886.
- [27] G. Kim, S. Lee, and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1690–1700, 2014. [Online]. Available: <https://doi.org/10.1016/j.eswa.2013.08.066>