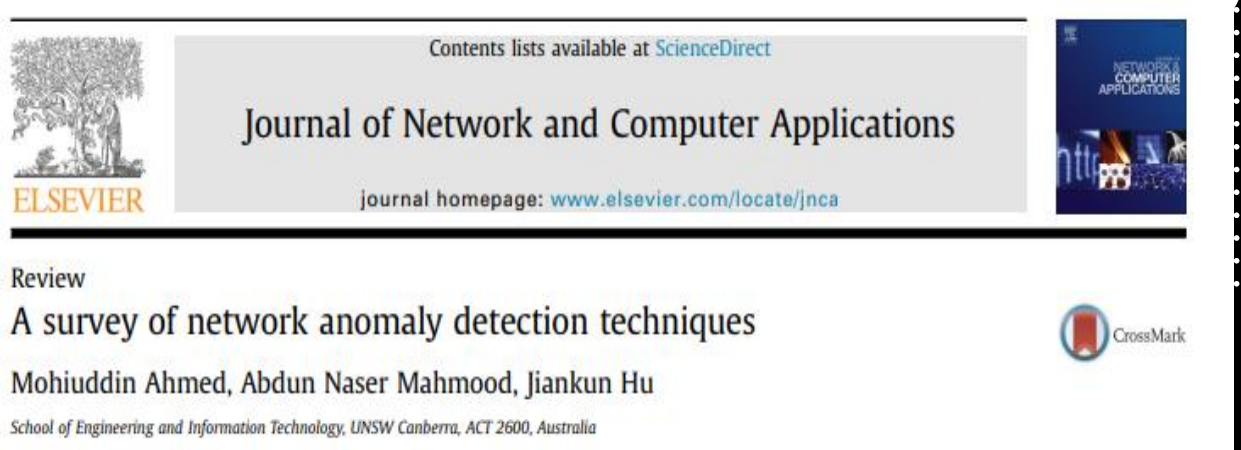
Network Anomaly Detection Techniques



- Amreen Batool
- AD202126001
- Deep Learning Applications A7602
- Professor: Yung-Cheol Byun

Table of Content

Introduction

- Anomaly detection is an important data analysis task that detects anomalous or abnormal data from a given dataset.
- Figure 1 shows the security incidents growth from 2009 to 2014. Therefore, the detection of network attacks has become the highest priority today. In addition, the expertise required to commit cyber crimes has decreased due to easily available tools

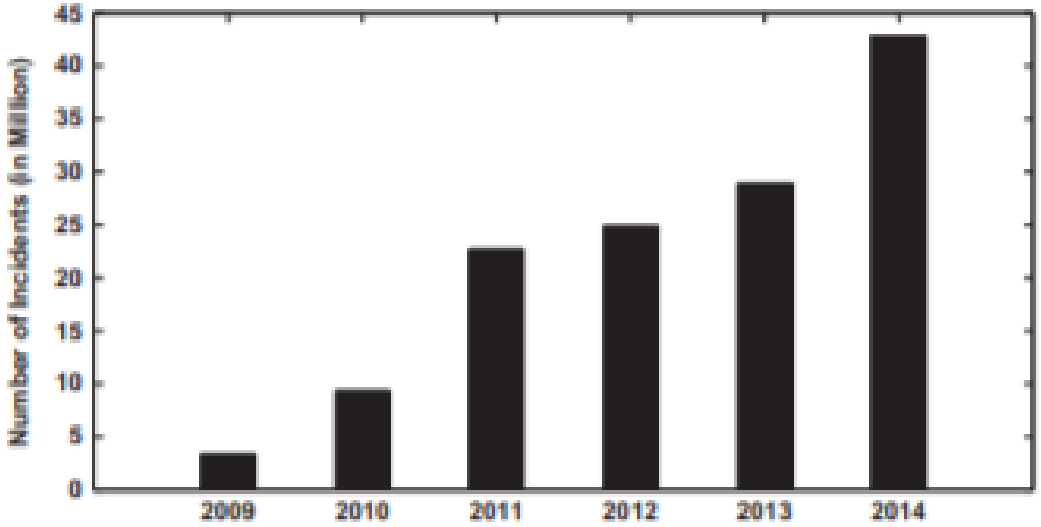


Fig. 1. Growth of information security incidents (The Global State of Information Security Survey, 2015).

Dataset

- Due to privacy issues, the datasets used for network traffic analysis are not easily available.
- There are very few publicly available datasets and among them DARPA/KDD datasets are considered as benchmark.
- KDD-CUP-99 Task Description (uci.edu)

Basic features of individual TCP connections.

feature name	*description*	*type*
duration	length (number of seconds) of the connection	continuous
protocol_type	type of the protocol, e.g. tcp, udp, etc.	discrete
service	network service on the destination, e.g., http, telnet, etc.	discrete
src_bytes	number of data bytes from source to destination	continuous
dst_bytes	number of data bytes from destination to source	continuous
flag	normal or error status of the connection	discrete
land	1 if connection is from/to the same host/port; 0 otherwise	discrete
wrong_fragment	number of ``wrong'' fragments	continuous
urgent	number of urgent packets	continuous

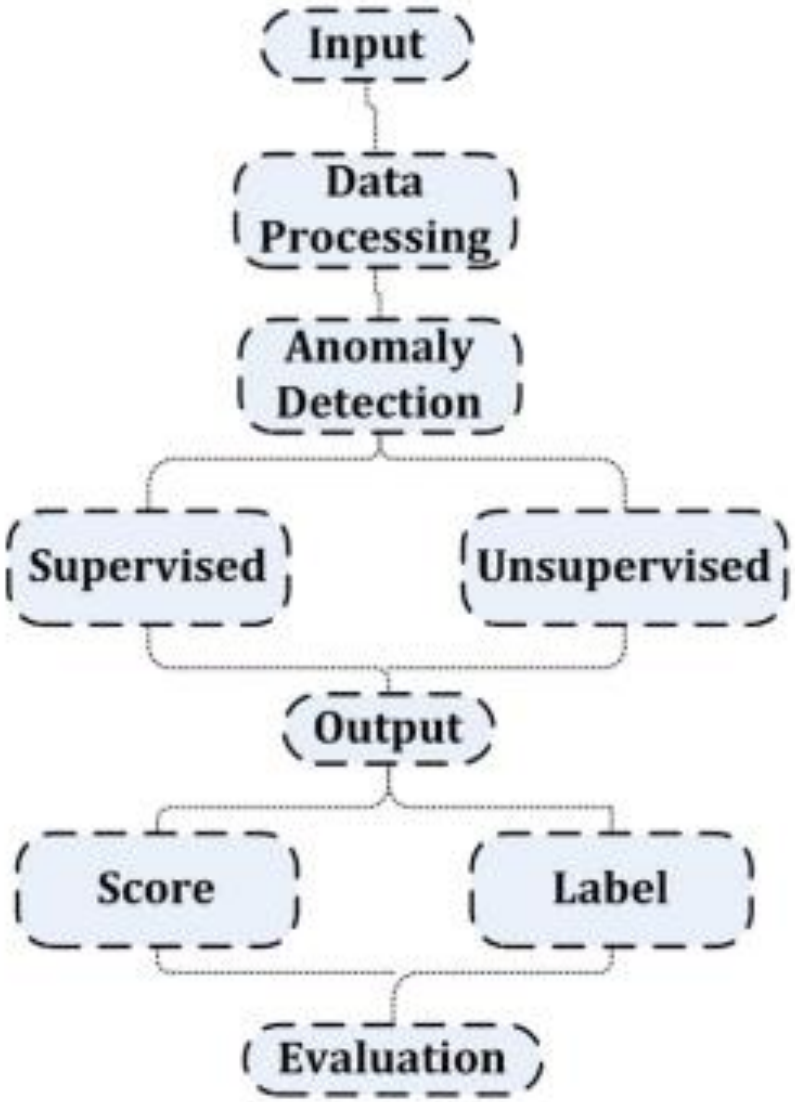
Content features within a connection suggested by domain knowledge.

feature name	*description*	*type*
duration	length (number of seconds) of the connection	continuous
protocol_type	type of the protocol, e.g. tcp, udp, etc.	discrete
service	network service on the destination, e.g., http, telnet, etc.	discrete
src_bytes	number of data bytes from source to destination	continuous
dst_bytes	number of data bytes from destination to source	continuous
flag	normal or error status of the connection	discrete
land	1 if connection is from/to the same host/port; 0 otherwise	discrete
wrong_fragment	number of ``wrong'' fragments	continuous
urgent	number of urgent packets	continuous

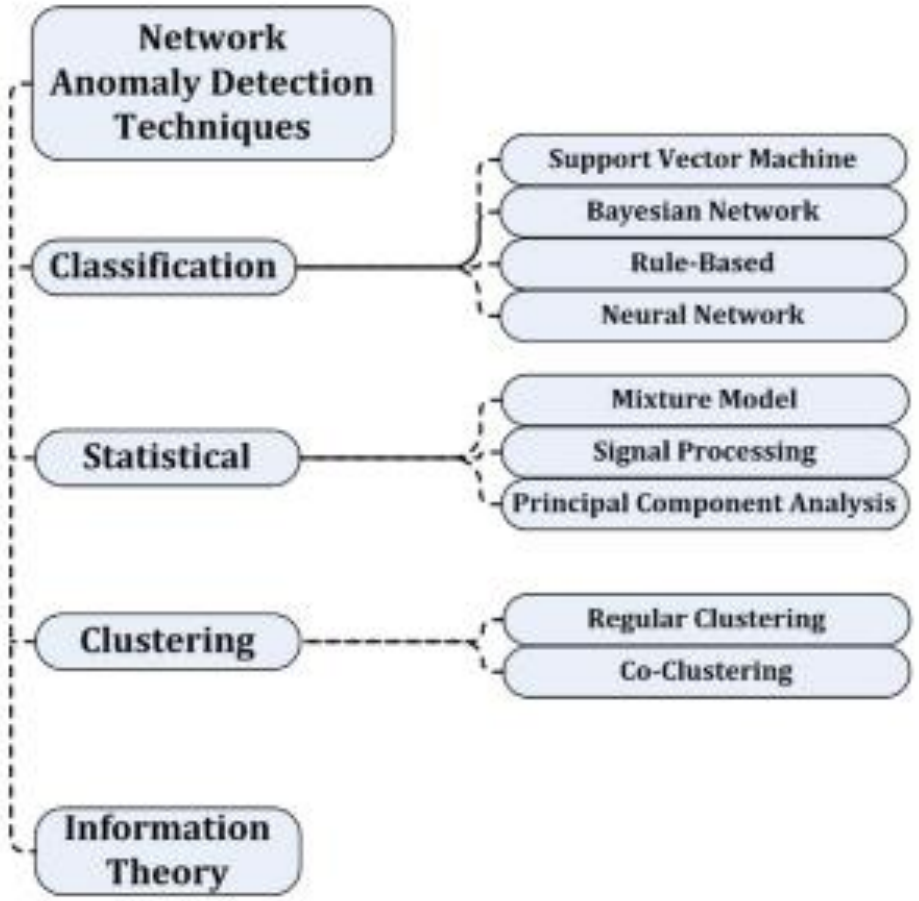
Traffic features computed using a two-second time window

feature name	*description*	*type*
count	number of connections to the same host as the current connection in the past two seconds	continuous
	Note: The following features refer to these same-host connections.	
serror_rate	% of connections that have ``SYN'' errors	continuous
rerror_rate	% of connections that have ``REJ'' errors	continuous
same_srv_rate	% of connections to the same service	continuous
diff_srv_rate	% of connections to different services	continuous
srv_count	number of connections to the same service as the current connection in the past two seconds	continuous
	Note: The following features refer to these same-service connections.	
srv_serror_rate	% of connections that have ``SYN'' errors	continuous
srv_rerror_rate	% of connections that have ``REJ'' errors	continuous
srv_diff_host_rate	% of connections to different hosts	continuous

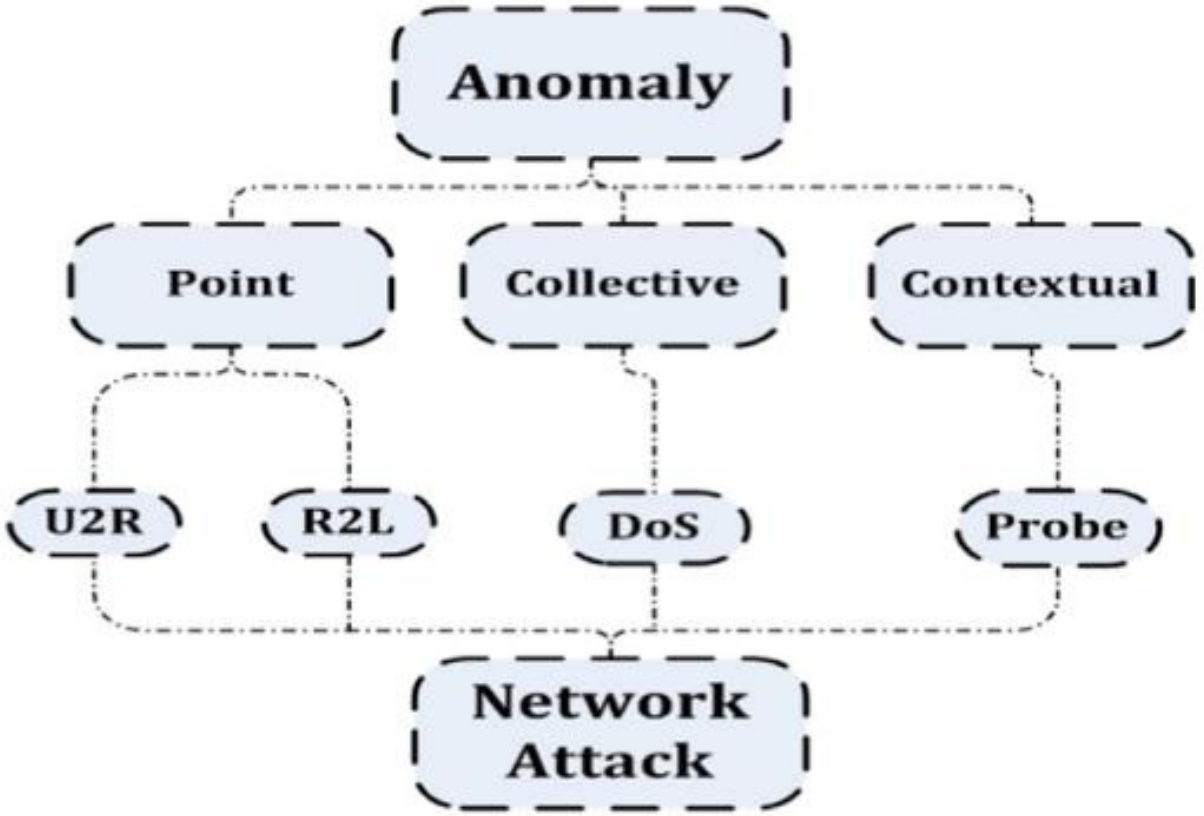
Generic framework for network anomaly detection

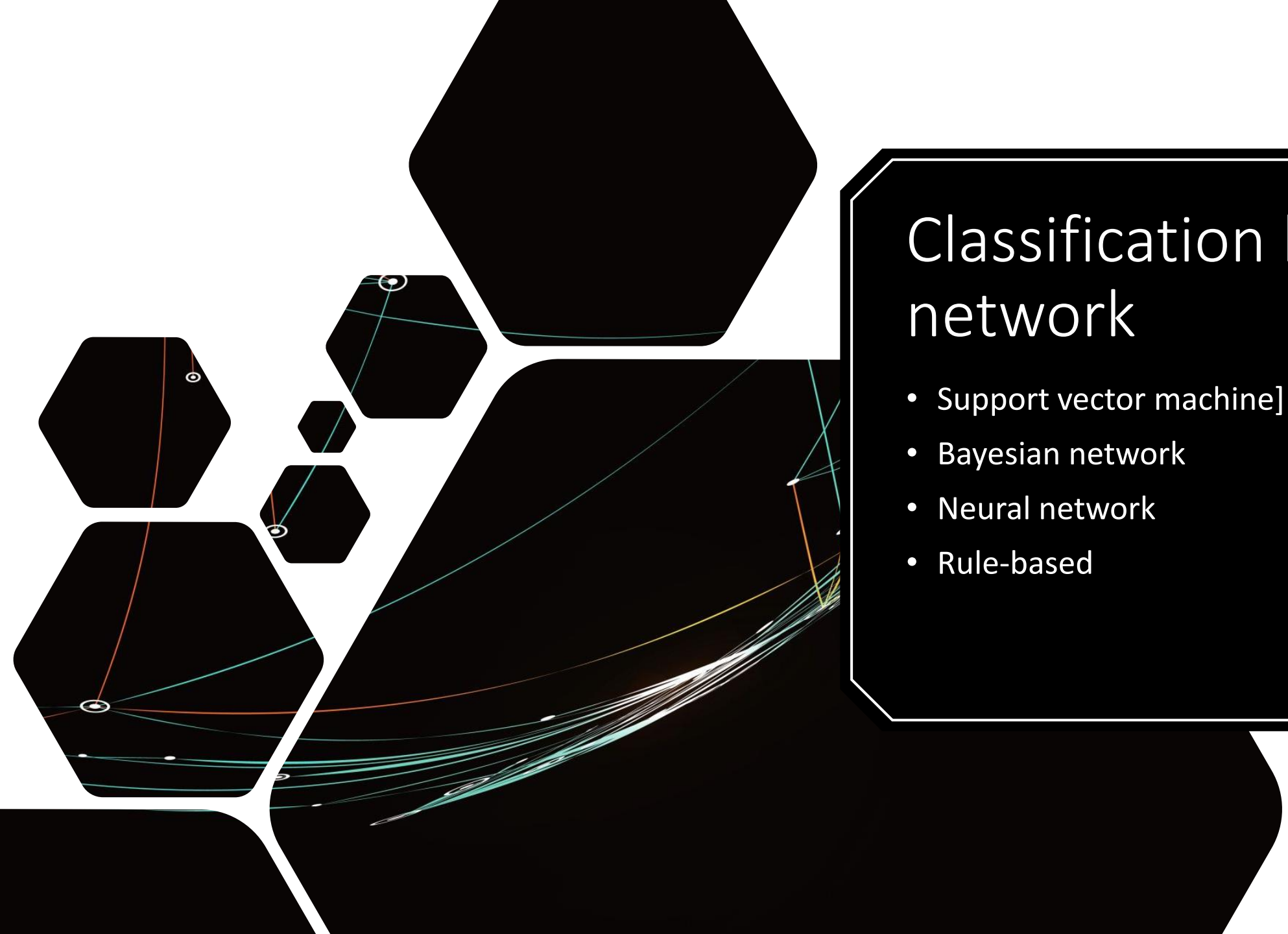


Taxonomy of network anomaly detection techniques



Classification based network





Classification based network

- Support vector machine]
- Bayesian network
- Neural network
- Rule-based

Support Vector Machine (SVM)

Support Vector Machine (SVM) is to derive a hyperplane that maximizes the separating margin between the positive and negative classes.

Bayesian network

A Bayesian network is an efficient approach for modeling a domain containing uncertainty.

$$\textbf{EC}(o_1, o_2, \ldots o_k 1)\ \boldsymbol{e\ is\ normal}\ \sum_{i=1}^{k} 0_i \leq I\ ||\boldsymbol{e\ is\ anomalous}$$

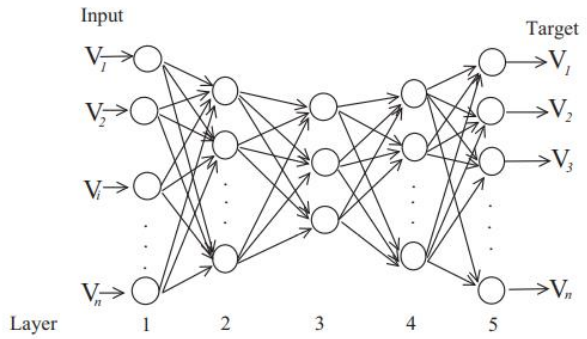
$$:\sum_{i=1}^{k} 0_i > I$$

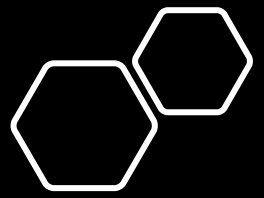
Neural network

Network anomaly detection, a neural network has been merged with other techniques, such as a statistical approach and variants of it

$$0=I_{ki}\sum_{j=0}^{lk-1} w_{kij} z_{(k-1)j}$$

where Iki is the weighted sum of the inputs to the unit, Zkj the output from the jth unit of the kth layer and Lk the number of units in the kth layer. The outlier factor is defined using the trained RNN as follows, where xij is the input value and oij the output value from the RNN





Rule-based

Rule-based anomaly detection techniques are widely used in supervised learning algorithms. These techniques consider both single and multi-label learning algorithms.

Statistical anomaly detection

- Intrusion detection techniques have also been developed using statistical theories; for example, the established chi-square theory is used for anomaly detection

$$X^2 = \sum_{i=1}^{n} \frac{(X_i - E_i)}{(E_i)}$$

- X_i= the observed value of the ith variable,
- E_i= the expected value of the iType equation here.th variable,
- n= the number of variables.
- Based on the principles of the statistical theory, different types of techniques have been developed to detect anomaly discussed next.

Information theory

- Information-theoretic measures can be used to create an appropriate anomaly detection model.
- Entropy

  $H(D) = \sum_{x \epsilon C_D} P(X) \log \frac{1}{P(x)}$
- Conditional Entropy

  $H(D|Y) = \sum_{xy \epsilon C_D, C_Y} P(x, y) \log \frac{1}{P(x|y)}$
- Relative entropy

  $(p|q) = \sum_{x \epsilon C_D} P(x) \log \frac{p(x)}{q(x)}$
- Relative conditional entropy

  $(p|q) = \sum_{x \epsilon C_D, C_Y} P(x, y) \log \frac{p(x|y)}{q(x|y)}$

Clustering-based

- Clustering refers to unsupervised learning algorithms which do not require pre-labeled data to extract rules for grouping similar data instances .
- Assumption 1: As we can create clusters of only normal data, any subsequent new data that do not fit well with existing clusters of normal data are considered anomalies.
- Assumption 2: When a cluster contains both normal and anomalous data, it has been found that the normal data lie close to the nearest clusters centroid, but anomalies are far away from centroids.
- Assumption 3: In a clustering with clusters of various sizes, the smaller and sparser can be considered anomalous and the thicker normal.

Result

Evaluation results.

Accuracy	Precision	Recall	F-measure	Attack cluster purity	Normal cluster purity
92.82%	0.9236	0.9923	0.96	92.36%	95.6%

Cluster purity comparison

Purity	Ahmed and Mahmood (2014b)	Papalexakis et al. (2012)
Normal (%)	95.6	75.84
Attack (%)	92.36	92.44

Evaluation of network anomaly detection techniques.

Technique	Output	Attack priority	Complexity
Classification	Label, score	DoS	Quadratic
Statistical	Label, score	R2L, U2R	Linear
Clustering	Label	DoS	Quadratic
Information theory	Label	Neutral	Exponential

Conclusions

In this research we described the assumptions for segregating normal data instances from anomalous. These assumptions will provide a guideline to assess the efficiency of the techniques when applied in a particular domain

Thank You