A Major Project II Report On

ANOMALY DETECTION TECHNIQUES

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

Submitted by:

Shreeya Mittal 2K18/CSE/15

Under the supervision of

Dr. Rajni Jindal

(Professor)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

OCTOBER, 2020

CANDIDATE'S DECLARATION

I, Shreeya Mittal , Roll No. 2K18/CSE/15 student of M.Tech (Computer Science Engineering), hereby declare that the project Dissertation titled **"Survey on Anomaly Detection Techniques"** which is submitted by me to the Department of Computer Science and Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of and Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi Shreeya Mittal

Date:

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering) Bawana Road, Delhi - 110042

CERTIFICATE

I hereby certify that the Project titled **"Survey on Anomaly Detection Techniques"** which is submitted by Shreeya Mittal, 2K18/CSE/15 Department of Computer Science Engineering, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere

Place: Delhi

Dr. Rajni Jindal

Date:

SUPERVISOR

Professor

ACKNOWLEDGEMENT

The success of a Major I project requires help and contribution from numerous individuals and the organization. Writing the report of this project work gives me an opportunity to express my gratitude to everyone who has helped in shaping up the outcome of the project.

I express my heartfelt gratitude to my project guide **Dr. Rajni Jindal** for giving me an opportunity to do my Minor I project work under his guidance. His constant support and encouragement has made me realize that it is the process of learning which weighs more than the end result. I am highly indebted to the panel faculties during all the progress evaluations for their guidance, constant supervision and for motivating me to complete my work. They helped me throughout by giving new ideas, providing necessary information and pushing me forward to complete the work.

I also reveal my thanks to all my classmates and my family for constant support.

Shreeya Mittal

ABSTRACT

In the present world huge amounts of data are stored and transferred from one location to another. The data when transferred or stored is primed exposed to attack. Although various techniques or applications are available to protect data, loopholes exist. The role of Intrusion Detection System (IDS) has been inevitable in the area of Information and Network Security – especially for building a good network defense infrastructure. Anomaly based intrusion detection technique is one of the building blocks of such a foundation. Thus to analyze data and to determine various kind of attack, data mining techniques have emerged to make it less vulnerable.

Anomaly detection uses these data mining techniques to detect the surprising behavior hidden within data increasing the chances of being intruded or attacked. Various hybrid approaches have also been made in order to detect known and unknown attacks more accurately. This paper presents a number of anomaly detection techniques that have been presented by various researchers.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Intrusion Detection

Intrusion, in simple terms, is an unlawful act of entering, seizing, or taking ownership of the property of another (the computer device being the property in this case). This means a code that disables the correct flow of traffic on the network or steals traffic information.

Intrusion detection in information security is the act of detecting acts that threaten to compromise a resource's confidentiality, integrity or availability. The computer based data processing era is recognized by the recent developments in the computer communication system. The need for intrusion detection stems from the fact that in each an d every aspect of this period of popular science and technology, computer systems are used . The following intrusion detection methodologies typically exist:—

1. Anomaly Detection -

It refers to detecting patterns that do not correspond to an existing standard behaviour in a given data set.The patterns thus observed are called anomalies and in many application domains also translate to important and actionable information. Often known as outliers, surprise, aberrant, anomaly, peculiarity, etc. are exceptions. In reality, it refers to storing the normal behaviour of users in the database, then comparing the actual behaviour of users with those in the database. If a significant enough divergence exists, it is said that the data checked is abnormal. The value of anomaly detection lies in its utter lack of device significance, its effective simplicity, and the potential to detect an attack that has never been detected before. Discrimination between malicious and legitimate patterns between operations (device or user driven) in variables characterising device normality is anomaly-based intrusion detection.

2. Misuse Detection

We first describe abnormal device behaviour in the misuse detection approach, and then describe all other behaviour as normal behaviour. It implies that there is an easy model to describe irregular conduct and behaviour. The benefit of incorporating proven attacks to the model is simplicity. Its downside is its failure to detect unknown attacks. Detection of Misuse refers to verifying instances of attack by matching features via the library of attacking features. This increases the high detection pace and the low number of false alarms. It fails, however, to discover the non-pre-designated attacks in the library of functions, so the numerous new attacks cannot be identified.

1.2 Challenges

An anomaly is characterized at an abstract level as a pattern that does not adhere to expected normal behavior. Therefore, a basic anomaly detection technique is to identify a region representing normal behavior and declare as an anomaly any finding in the data that does not belong to this normal region. But many variables make this seemingly straightforward solution very difficult:

i. It is very difficult to describe a normal region that includes all conceivable normal activity. Furthermore, the boundary between natural and abnormal conduct is also not precise. An anomalous observation close to the boundary may therefore actually be natural, and vice versa.

ii. If anomalies are the product of malicious acts, malicious adversaries often change to make the anomalous observations appear natural, making it more difficult to identify natural conduct. Normal behavior continues to develop in several domains and a current definition of normal behavior may not be adequately reflective in the future.

iii. For various application domains, the exact notion of an anomaly is various For example, a minor deviation from normal in the medical domain (e.g., body temperature fluctuations) might be an anomaly, whereas a similar deviation in the stock market

domain (.g., stock value fluctuations) could be considered normal. It is therefore not straightforward to adapt a technique built in one domain to another.

iv. The availability of labelled data for training / validation of models used by techniques for anomaly detection is typically a major problem.

v. The knowledge also includes noise that appears to be identical to the real anomalies and is thus hard to differentiate and eliminate.

Due to the above challenges, the anomaly detection problem, in its most general form, is not easy to solve. In fact, most of the existing anomaly detection techniques solve a specific formulation of the problem. The formulation is induced by various factors such as nature of the data, availability of labeled data, and type of anomalies to be detected, etc. Often, these factors are determined by the application domain in which the anomalies need to be detected. Researchers have adopted concepts from diverse disciplines such as statistics, machine learning, data mining, information theory, spectral theory, and have applied them to specific problem formulations.

2.7Data Mining

The method of extracting patterns from data is data mining. Data mining is seen by modern business as an increasingly valuable method to turn data into business intelligence, offering an knowledge advantage. Currently, it is used in a wide variety of profiling activities, such as marketing, tracking, identification of fraud, and scientific exploration. A primary purpose for using data mining is to better interpret samples of behavioural observations. An inevitable fact of data mining is that the (sub-) set(s) of data being analysed may not be representative of the entire domain and may therefore not provide examples of such important relationships and behaviours that occur in other parts of the domain.

Technology is advanced in data mining for:

i. Hidden and neglected knowledge can be found.
ii. It can process a vast amount of data.

Four types of tasks are generally involved in data mining:—

i. Clustering-It is the task of identifying groups and structures in the data that are "similar" in some way or another, without using known data structures.

ii. Classification-it is the process of generalising the known structure for new data to be implemented. An email programme, for instance, might try to identify an email as legitimate or spam. Decision tree learning, nearest neighbour, Naive Bayesian classification, neural networks, and support vector machines are typical algorithms.

iii. Regression- given a specific dataset, predicts a number of numeric values (also called continuous values). Regression, for example, may be used, given other variables, to estimate the cost of a product or service.

iv. Association rule learning -Looks for associations between variables. For example, a supermarket may collect information on shopping patterns for customers. The supermarket will decide which items are often purchased together, using association rule learning, and use this knowledge for marketing purposes. This is often referred to as analyzing consumer baskets.

CHAPTER 2

ANOMALY DETECTION

2.1 Different aspects of anomaly detection

The various aspects of anomaly detection are described and discussed in this section. As stated earlier, several different variables, such as the existence of the input data, the availability (or unavailability) of labels, as well as the constraints and requirements caused by the application domain, decide the particular formulation of the problem. This section illustrates the richness of the problem domain and justifies the need for a wide variety of techniques for detecting anomalies.

2.1.1 Nature of Input Data

The essence of the input data is a core feature of any anomaly detection technique. Input is typically a set of instances of data (also referred to as object, record, point, vector, pattern, event, case, sample, person, observation). A collection of attributes (also referred to as vector, characteristic, function, area, dimension) can be used to define any data case. There may be various types of attributes, such as binary, categorical or continuous. Each instance of data could consist of either one (univariate) or multiple (multivariate) attributes. Both attributes may be of the same type in the case of multivariate data cases, or may be a mixture of different data types.

The existence of attributes defines the applicability of techniques for anomaly detection. For example, different statistical models have to be used for continuous and categorical data in the case of statistical techniques. Similarly, the existence of attributes will decide the distance measure to be used for nearest neighbour dependent approaches. Sometimes, the pair distance between instances could be given in the form of a distance (or similarity) matrix

instead of the actual data. In such cases, many statistical and classification-based techniques are not applicable to techniques requiring original data instances, e.g. based on the relationship between data cases, input data may also be classified. Most of the current techniques of anomaly detection deal with record data (or point data) in which no relationship between data instances is assumed.

Data instances may be connected to each other in general. Sequence data, spatial data, and data from graphs are some examples. The data instances are linearly ordered in sequence data, e.g., time series data, genome sequences, protein sequences. Each data instance is connected to its neighbouring instances in spatial data, e.g., vehicular traffic data, ecological data. If there is a temporal (sequential) aspect of the spatial data, it is referred to as spatio-temporal data, such as climate data. Data instances are represented as vertices in a graph in graph data and are associated with other vertices with edges. Later in this chapter, we will address situations where the relationship between data instances becomes important to the detection of anomalies.

2.1.2 Type of Anomaly

The essence of the desired anomaly is an important feature of an anomaly detection technique. In the following three groups, anomalies can be classified:

i. *Point Anomalies*. If, in comparison to the rest of the data, an individual data instance can be considered anomalous, then the instance is called a point anomaly. This is the simplest kind of anomaly and is the subject of the majority of anomaly detection studies.

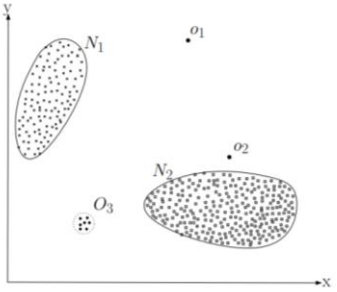


Figure 1

For example, points o1 and o2 as well as points in region O3 are located beyond the limits of the normal regions in Figure 1, and are thus point anomalies as they vary from the normal data points.

Consider credit card fraud detection as a real life example. Let the data set correspond to the credit card transactions of a person. Let us presume, for the sake of simplicity, that the data is described using only one function: the amount spent. A point anomaly is a transaction for which the amount spent is very high compared to the usual expenditure range for that person.

ii. *Contextual Anomalies*. If, in a particular sense, a data instance is anomalous (but not otherwise), it is considered a contextual anomaly (also known as a conditional anomaly. The notion of a context is induced by the data set structure and must be defined as part of the problem formulation. The following two sets of attributes are used to describe each data instance:

(1) Contextual Attributes. For this case, contextual attributes are used to determine the context (or neighbourhood). For starters, the longitude and latitude of a position are the contextual attributes in spatial data sets. Time is a contextual attribute in time series data that specifies an instance's location on the whole list.
(2) Behavioral attributes. The behavioural attributes characterize an instance's non-contextual features. For instance, the amount of rainfall at any location is a behavioral attribute in a spatial data set representing the average rainfall of the entire planet.

The anomalous behaviour is calculated within a given context using the values for the behavioural attributes. In a given context, a data instance may be a contextual anomaly, but in a different context, an equivalent data instance (in terms of behavioural attributes) might be considered natural. For a contextual anomaly detection technique, this property is important in defining contextual and behavioural attributes.

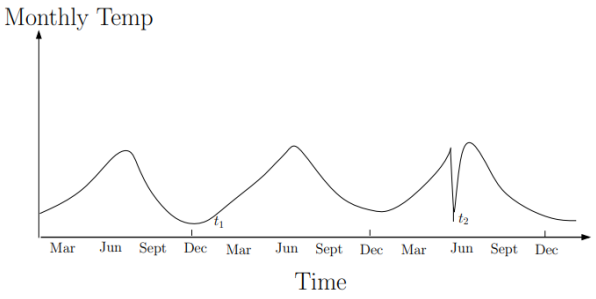


Figure 2

In time series data and spatial data, contextual anomalies were most frequently investigated. One such example of a temperature time series showing the monthly temperature of a region over the last few years is shown in Figure 2. A 35F temperature could be common at that location during the winter (at time t1), but the same value would be an anomaly during the summer (at time t2).

In the credit card fraud detection area, a similar example can be found. The time of acquisition may be a qualitative feature in the credit card domain. Suppose a person normally has a $100 weekly shopping bill, except when it exceeds $1000 during Christmas week. A new purchase of $1000 per week in July would be deemed a contextual anomaly as it does not correspond to the individual's usual actions in the context of time (although it will be considered normal to spend the same amount during Christmas week).

The meaningfulness of the contextual anomalies in the target application domain dictates the option of applying a contextual anomaly detection technique. The availability of contextual attributes is another key factor. It is easy to identify a context in many situations, and it makes sense to apply a contextual anomaly detection technique. In other instances, it's not easy to identify a context, making it difficult to implement those techniques.

iii. Collective Anomalies.

If, with respect to the entire data set, a series of associated data instances is anomalous, it is called a collective anomaly. In a collective anomaly, the individual data instances may not be anomalies individually, but their occurrence together as a collection is anomalous. An example that shows a human electrocardiogram output is illustrated in Figure 3. As the same low value occurs for an abnormally long period (corresponding to an Atrial Premature Contraction), the highlighted area denotes an anomaly. Notice that, by itself, that low value is not an anomaly.

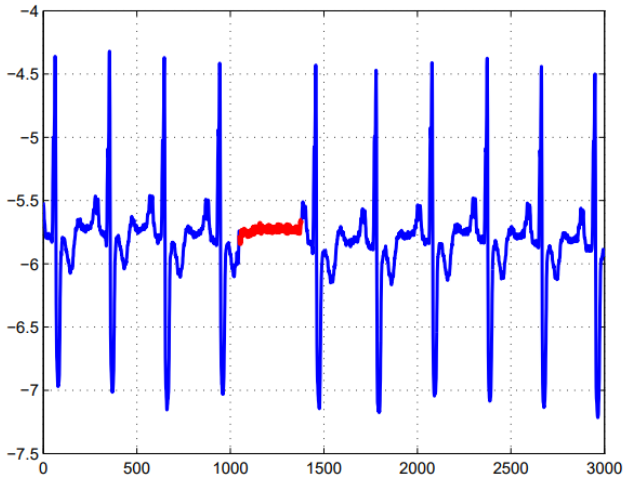


Figure 3

It should be noted that while point anomalies can occur in any data set, only in data sets in which data instances are connected can collective anomalies occur. Contextual anomalies, on the other hand, depend on the availability of context attributes in the data. If evaluated relative to a context, a point anomaly or a collective anomaly may also be a contextual anomaly. Thus, by integrating the background information, a point anomaly detection problem or collective anomaly detection problem can be converted to a contextual anomaly detection problem.

2.1.3 Data Labels

The labels associated with an instance of data indicate whether that instance is regular or anomalous. It should be noted that it is always prohibitively costly to procure branded data that is reliable as well as descriptive of all forms of behaviours. Labeling is often performed by a human expert manually, and thus requires considerable effort to acquire the labelled training data collection. Usually, it is more difficult to get a classified set of anomalous data instances representing all possible forms of anomalous behaviour than to get labels for normal behaviour. In addition, anomalous behaviour is also complex in nature, e.g. new forms of anomalies may occur, for which no training data is labelled. Anomalous incidents would translate into disastrous occurrences in some situations, such as air traffic safety, and would therefore be extremely rare.

Anomaly detection techniques can work in one of the following three modes, depending on the degree to which the labels are available:

i. **Supervised Detection of Anomaly:** Techniques trained in supervised mode presume that a training data set that has classified instances for both normal and anomaly groups is available. For normal vs. anomaly classes, a common approach in such cases is to construct a predictive model. To decide the class it belongs to, every unseen data instance is compared against the standard. In supervised anomaly detection, there are two big problems that occur. First, relative to the usual cases in the training results, the anomalous instances are much smaller. Second, it is generally difficult to obtain reliable and representative labels, in particular for the anomaly class.

ii. **Semi-supervised Detection of anomalies:** Techniques working in a semi-supervised mode presume that only the usual class has labelled instances with the training data. They are more commonly applicable than supervised techniques because they do not need labels for the anomaly class. For instance, an anomaly scenario in spacecraft fault detection will mean an accident that is not easy to model. Building a model for the class corresponding to normal behaviour and using the model to classify anomalies in the test data is the standard approach used in such techniques.

iii. **Detection of unsupervised anomalies**: Techniques operating in unsupervised mode do not require, and are thus most commonly applicable, training data. The methods in this category allow the implicit assumption that regular instances are much more common in the test data than anomalies. These techniques suffer from a high false alarm rate if this statement is not valid.

2.1.4 Output of Anomaly Detection

For any anomaly detection method, the way in which the anomalies are identified is an significant feature. Usually, one of the following two forms is the outputs provided by anomaly detection techniques:

i. *Scoring*. In the test results, scoring techniques give each instance an anomaly score based on the degree to which that instance is considered an anomaly. The performance of such methods is therefore a categorised list of anomalies. An analyst can choose to either evaluate the top few anomalies or choose to select the anomalies using a cut-off threshold.

ii. *Labels*. Techniques in this category assign each test instance to a mark (normal or anomalous). Scoring-based methods of anomaly detection allow the analyst to select the most significant anomalies using a domain-specific threshold. Techniques that provide the test cases with binary labels do not allow analysts to make such a choice explicitly, although this can be managed indirectly within each technique by parameter choices.

2.2 Applications of Anomaly Detection

We address many applications of anomaly detection in this chapter. The following four aspects are discussed for each application domain:

i. The concept of anomaly.
ii. Nature of data.
iii. Challenges connected with anomaly detection.
iv. Current methods for anomaly detection.

2.2.1 Intrusion Detection

Intrusion detection refers to the detection in a computer-related device of malicious behaviour (break-ins, penetration, and other types of computer abuse. From a computer security perspective, these malicious behaviours or intrusions are interesting. An intrusion is different from the system's usual behaviour, and anomaly detection techniques are also applicable in the area of intrusion detection.

The huge volume of data is the key challenge for anomaly detection in this domain. In order

to manage these large-sized inputs, anomaly detection techniques need to be computer-efficient. In addition, the information usually comes in a streaming mode, requiring on-line review. The false alarm rate is another problem that occurs because of the large input size. A few percent of false alarms can make research daunting for an analyst, because the data amounts to millions of data items. Labeled data corresponding to normal behaviour is commonly available, while intrusion labels are not. Thus, in this domain, semi-supervised and unsupervised anomaly detection techniques are favoured.

Intrusion detection systems are usually categorised into **host-based** and **network-based** intrusion detection systems.

i. Host Based Intrusion Detection Systems.

These systems (also referred to as system call detection systems for intrusion) deal with call traces of the operating system. The intrusions take the form of the traces' anomalous subsequences (collective anomalies). Anomalous subsequences are translated into malicious programmes, improper actions, and breaches of policy. While all traces include events belonging to the same alphabet, the main element in distinguishing between regular and anomalous activity is the co-occurrence of events.

To handle the sequential nature of data, anomaly detection techniques applied to host-based intrusion detection are needed. In addition, in this domain, point anomaly detection techniques are not available. The methods must either model the data of the sequence or compute the similarity between sequences.

ii. Network Intrusion Detection Systems.

Such systems deal with detecting network data intrusions. Usually, the intrusions occur as anomalous patterns (point anomalies), while some methods sequentially model the data and detect anomalous subsequences (collective anomalies). The primary explanation for these anomalies is due to the attacks undertaken for data theft or destruction of the network by outside hackers who want to obtain unauthorised access to the network. A typical environment is a large computer network which is connected via the Internet to the rest of the world.

The usable data for intrusion detection systems can be at various granularity levels, e.g. traces of packet level, CISCO net-flow data, etc. The data has a temporal aspect associated with it, but the sequential aspect is usually not handled directly by most techniques. Typically, the data is high-dimensional with a combination of both categorical and continuous attributes.
A challenge faced in this area by anomaly detection techniques is that the existence of anomalies continues to evolve over time as intruders adapt their network attacks to avoid current solutions for intrusion detection. In Table I, some anomaly detection techniques used in this domain are mentioned.

TECHNIQUE USED	REFERENCES
Bayesian Network	Siaterlis and Maglaris [2004]
Neural Networks	Ramadas et al. [2003]
Nearest Neighbor Based	Chandola et al. 2006]

Table I

2.2.2 Fraud Detection

Detection of fraud refers to the detection of fraudulent activity in commercial organisations, such as banks, credit card firms , insurance companies, mobile phone companies, stock markets, etc. The malicious users may be the organisation's own clients or may pose as a client (also known as identity theft). Fraud occurs when the services offered by the company are consumed by these users in an unauthorised manner. In order to avoid economic losses, organisations are involved in the prompt identification of such fraud.
The word activity monitoring is presented by Fawcett and Provost[1999] as a general approach to fraud detection in these domains. Maintaining a usage profile for each customer and tracking the profiles to detect any anomalies are the standard method of anomaly detection techniques. Some of the specific fraud detection applications are discussed below.

i. Credit Card Fraud Detection.

Anomaly detection techniques are used in this domain to identify fraudulent credit card applications or fraudulent use of credit cards (associated with theft of credit cards). Close to detecting insurance fraud is the identification of fraudulent credit card applications.

The data usually consists of multi-dimensional records such as the user ID, sum spent, time between the use of consecutive cards, etc. Usually, the frauds are expressed in transactional records (point anomalies) and lead to high purchases, purchase of products never bought by the consumer before, high purchase volume, etc. The credit agencies have full details available and have documents labelled as well. Moreover, depending on the credit card customer, the data falls into different profiles. Therefore, techniques based on profiling and clustering are usually used in this domain. The difficulty associated with detecting illegal use of credit cards is that, as soon as the fraudulent transaction takes place, it needs online fraud detection.

To resolve this problem, anomaly detection techniques have been implemented in two separate ways. The first one is classified as a by-owner in which, based on its credit card use history, each credit card user is profiled. Every new transaction is matched with the profile of the customer and flagged as an anomaly if the profile does not fit. This method is usually costly, because any time a user makes a transaction, it involves querying a central data repository. Another method, known as by-operation, describes anomalies that arise at a certain geographical location between transactions. The contextual irregularities are observed by both the by-user and by-operation techniques. The context is a user in the first case, while the context is a geographic place in the second case. In Table II, some anomaly detection techniques used in this domain are mentioned.

TECHNIQUE USED	**REFERENCES**
Rule Based	Bolton and Hand [1999]
Neural Networks	Brause et al. [1999]

Table II

ii. Insurance Claim Fraud Detection.

Allegations of fraud, such as car insurance fraud, are an important concern in the property insurance industry. For unauthorized and unlawful claims, individuals and conspiratorial

rings of claimants and providers exploit the claim processing method. For the related firms, the detection of such fraud was very necessary in order to prevent financial losses.

Documents submitted by the claimants are the available data in that domain. From these records, the methods extract various features (both categorical and continuous). Claim adjusters and investigators usually investigate these claims for fraud. These manually examined cases are used by supervised and semi-supervised techniques for insurance fraud identification as labelled instances.

As a generic behaviour tracking issue, insurance claim fraud detection is very frequently viewed. Techniques based on the neural network were also used to classify anomalous insurance claims.

iii. Mobile Phone Fraud Detection.

A typical activity tracking problem is mobile / cellular fraud detection. The task is to search a vast range of accounts, analyse each one's calling activity, and issue a warning when it appears that an account has been misused.

Calling activities may be interpreted in different ways, but call records are commonly defined. Each call record is a continuous (e.g., CALL-DURATION) and discrete (e.g., CALLING-CITY) function vector. In this domain, however, there is no inherent primitive representation. Depending on the granularity required, calls may be aggregated by time, for example into call hours or call days or user or location. The anomalies reflect the high volume of calls to unexpected destinations or calls made. In Table III, some anomaly detection techniques used in this domain are mentioned.

TECHNIQUE USED	**REFERENCES**
Rule Based	Phua et al. [2004],Taniguchi et al. [1998]
Neural Networks	Taniguchi et al. [1998]
Histogram Based	Fawcett and Provost [1999]

Table III

2.2.3 Medical and Public Health Anomaly Detection

Anomaly identification usually deals with patient records in the medical and public health realms. Due to different causes, such as irregular patient condition or instrumentation failures or recording failures, the data may have anomalies. The identification of disease outbreaks in a geographic region has also been based on many techniques. The identification of anomalies is therefore a very important issue in this domain and requires a high degree of precision.

Typically, the data consists of records that may have several different forms of characteristics, such as patient age, weight, blood group. There may also be temporal as well as spatial aspects to the data. Many of this domain's existing anomaly detection techniques aim to detect anomalous records (point anomalies). The labelled data usually belongs to healthy patients, so most of the approaches follow a semi-supervised approach. Time series data, such as electrocardiograms (ECG) (Figure 3) and electroencephalograms (EEG), is another type of data treated by anomaly detection techniques in this domain. To detect anomalies in such data, collective anomaly detection techniques have been applied. The most difficult aspect of this domain's anomaly detection problem is that the cost of classifying an anomaly as normal can be very high. In Table IV, some anomaly detection techniques used in this domain are mentioned.

TECHNIQUE USED	REFERENCES
Rule Based	Aggarwal [2005]
Nearest Neighbor Based	Lin et al. [2005]

Table IV

2.2.4 Industrial Damage Detection

Because of constant use and the ordinary wear and tear, industrial units suffer damage. In order to avoid further escalation and damages, such damage needs to be detected early. Typically, the data in this domain is referred to as sensor data since it is registered and collected for analysis using various sensors. To detect such injury, anomaly detection

techniques have been extensively used in this domain. It is possible to further categorize industrial damage detection into two domains, one dealing with defects in mechanical components such as motors, motors, etc., and the other dealing with defects in physical structures. The former domain is often referred to as management of machine health.

i. Fault Detection in Mechanical Units.

In this domain, anomaly detection techniques monitor the output of industrial components such as generators, turbines, pipeline oil flow or other mechanical components and detect defects that may occur due to wear and tear or other unexpected circumstances.
Usually, the data in this domain has a temporary aspect and some techniques often use time series analysis. The anomalies arise mainly in a single context (contextual anomalies) or as an anomalous series of observations (collective anomalies) because of an observation.
Standard data (related to components without defects) is usually readily available and semi-supervised techniques are also applicable. Anomalies must be reported in an online manner, as preventive steps must be taken as soon as an anomaly occurs.
In Table V, some anomaly detection techniques used in this domain are mentioned.

TECHNIQUE USED	**REFERENCES**
Rule Based	Yairi et al. [2001]
Neural Networks	Diaz and Hollmen [2002]

Table V

ii. Structural Defect Detection.

Techniques for structural defects and damage identification detect structural abnormalities in structures, e.g. beam fractures, airframe strains.
There is a temporal component to the data obtained in this domain. The techniques of anomaly detection are similar to techniques of novelty detection or change point detection because they seek to detect changes in the data obtained from a structure. Over time, the

normal data and therefore the models learned are usually static. There may be spatial correlations in the data. In Table IX, some anomaly detection techniques used in this domain are mentioned.

2.2.5 Anomaly Detection in Text Data

In a series of documents or news articles, anomaly detection techniques in this area mainly detect novel topics or incidents or news reports. Due to a new interesting event or an anomalous subject, the anomalies are caused. Typically, the data in this domain is strongly dimensioned and very sparse. Because the records are gathered over time, the data also has a temporal component. A challenge in this domain for anomaly detection techniques is to deal with the wide variations in documents belonging to one category or subject. Table VI lists several anomaly detection techniques used in this domain.

TECHNIQUE USED	REFERENCES
Clustering Based	Srivastava [2006]
Neural Networks	Manevitz and Yousef [2000]

Table VI

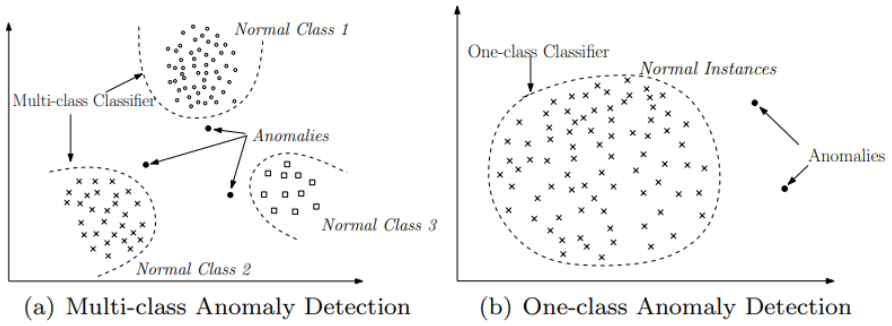
2.3 Classification based Anomaly Detection

Classification is used to learn a model (classifier) from a series of instances of labelled data (training) and then classify a test instance using the learned model (testing) into one of the classes. Anomaly detection strategies based on classification work in a similar two-phase fashion. Using the available labelled training data, the training stage learns a classifier. Using the classifier, the testing stage classifies a test instance as regular or anomalous.

Anomaly detection strategies based on classification work under the following general assumption:

Assumption: In the given feature space, a classifier which can differentiate between usual and anomalous classes can be taught.

Classification based anomaly detection techniques can be divided into two broad groups based on the labels available for the training phase: multi-class and one-class anomaly detection techniques. Multi-class anomaly detection techniques based on classification presume that labelled instances belonging to several normal classes are found in the training data. Such methods of anomaly detection teach a classifier to differentiate between each regular class and the rest of the classes. For diagrams, see Figure 4(a). A test example is considered to be anomalous if all of the classifiers do not classify it as natural. Some methods in this sub-category compare the prediction made by the classifier with a confidence score. The instance is declared to be anomalous if none of the classifiers are secure in classifying the test instance as usual.



(a) Multi-class Anomaly Detection (b) One-class Anomaly Detection

Figure 4

One-class anomaly detection strategies focused on classification presume that all training instances have only one class mark. Such techniques use a one-class classification algorithm to learn a discriminatory boundary across normal instances, e.g. one-class SVMs, one-class Kernel Fisher Discriminants, as shown in Figure 4(b). Any instance of the test that does not fall within the boundary learned is declared anomalous.

2.3.1 Bayesian Network Based

There are several situations in which there are statistical dependencies or causal associations between device variables. The probabilistic relationships between these variables can be difficult to precisely express. In other words, the previous interpretation of the method is simply that others may be affected by some variable.

A probabilistic graph model, called Naïve Bayesian Networks (NB), can be used to take advantage of this structural relationship between the random variables of a problem. This model answers questions such as what the likelihood of a certain form of attack is if few observed events are given. It can be achieved by using the conditional probability formula. A Directed Acyclic Graph (DAG) usually represents the structure of an NB, where each node represents one of the system variables and each relation encodes the effect of one node on another. When decision tree and Bayesian techniques are compared, decision tree accuracy is much better, but Bayesian network computational time is poor. Therefore, it is effective to use NB models when the data set is very large.

For network intrusion detection, novelty detection in video surveillance, anomaly detection in text data and disease outbreak detection, many variants of the basic technique have been suggested. The basic technique mentioned above assumes that the various attributes are independent. It has been suggested that many variants of the basic technique capture the conditional dependencies between the various attributes using more complex Bayesian networks.

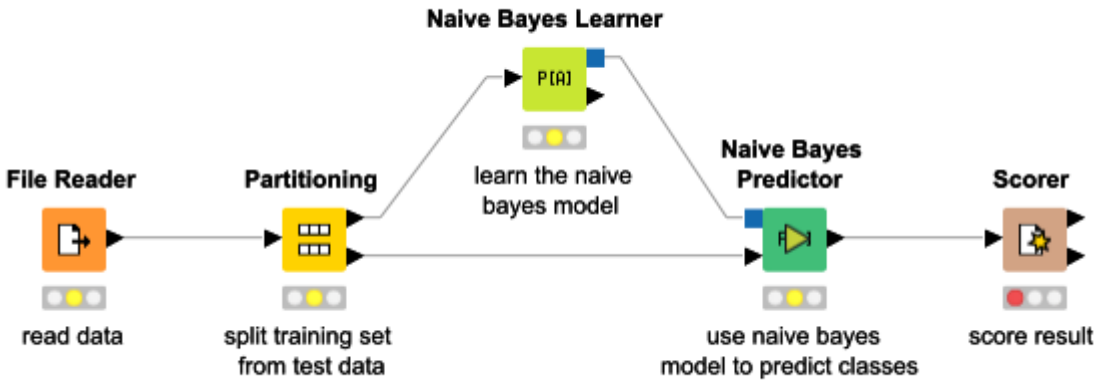


Figure 5

2.3.2 Neural Network Based

It is a series of interconnected nodes that are programmed to mimic the human brain's functioning. There is a weighted relation between each node and several other nodes in

adjacent layers. The input obtained from connected nodes is taken by individual nodes and the weights are used along with a simple function to calculate output values. For supervised or unsupervised learning, neural networks can be built. The user determines both the number of hidden layers and the number of nodes within the hidden layer in question. The output layer of the neural network can contain one or several nodes, depending on the application.

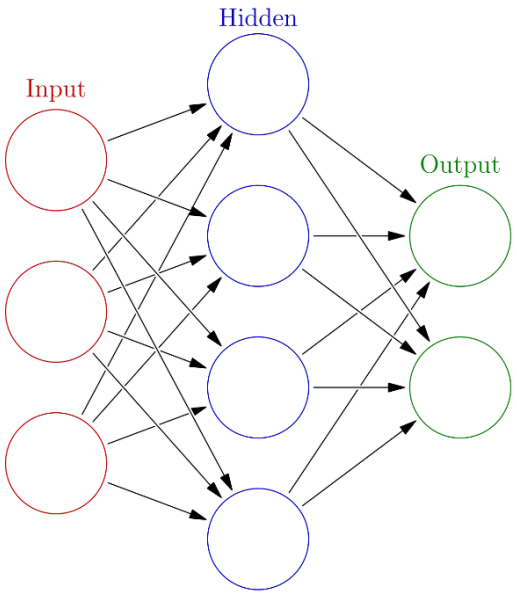


Figure 6

2.3.3 Support Vector Machine

These are a set of similar techniques of supervised learning used for classification and regression. The field of pattern recognition is commonly applied to the Support Vector Machine (SVM).

An SVM model is a representation of the examples as points in space, mapped such that a simple distance that is as large as possible separates the examples of the individual categories. In addition to performing linear classification, a non-linear classification can be effectively done by SVMs, implicitly mapping their inputs into high-dimensional feature spaces. A model that assigns new examples to one category or another is generated by the SVM training algorithm, making it a non-probabilistic binary linear classifier. A Support Vector Machine (SVM) is a formally defined hyperplane-separating discriminative classifier. In other words, the algorithm outputs an optimal hyperplane that categorises new

instances, given labelled training data (supervised learning).

It is also used for the detection method of an intrusion. As compared to neural networks in the KDD cup data collection, it was found that in most types of attacks, SVM performed NN in terms of false alarm rate and precision.

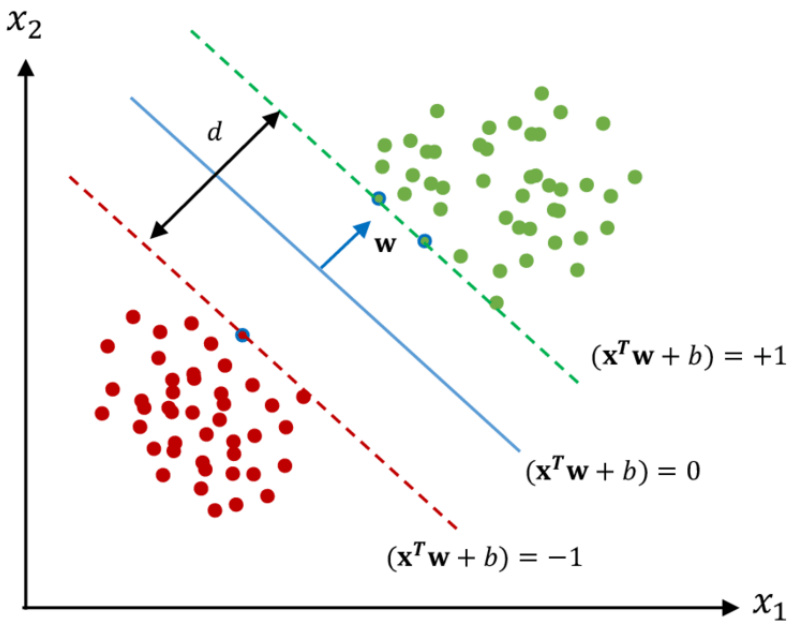


Figure 7

2.3.4 Rule Based

Rule-based anomaly detection methods learn laws that capture a system's normal behaviour. An example of a test that is not protected by any such law is known as an exception. In multi-class and one-class environments, rule-based techniques have been applied.

There are two steps to a simple multi-class rule based technique. The first step is to learn rules using a rule learning algorithm from training data, such as RIPPER, Decision Trees, etc. Each rule has an associated trust value that is proportional to the ratio of the number of training instances properly classified by the rule to the total number of training instances protected by the rule. The second step is to find the rule that best captures the test example for each test case. The inverse of the trust associated with the best law is the test instance's anomaly score.

Advantages and Disadvantages of Classification Based Techniques

The benefits of approaches focused on classification are as follows:

i. Classification-based techniques can make use of powerful algorithms that can differentiate between instances belonging to different classes , especially multi-class techniques.
ii. As each test instance needs to be compared against the pre-computed model, the testing process of classification-based techniques is rapid.

The drawbacks of strategies based on classification are as follows:

i. Strategies focused on multi-class classification depend on the availability of precise labels for different standard classes, which is often not possible.
ii. Classification-based techniques give each test instance a label, which can also become a drawback if the test cases need a significant anomaly score. To tackle this problem, some classification techniques that obtain a probabilistic prediction score from the performance of a classifier can be used.

2.4 Clustering based Anomaly Detection

It is possible to describe clustering as the splitting of data into groups of similar objects. Each group, or cluster, consists of objects that are similar to each other and different to other groups of objects. Clustering algorithms will, without prior knowledge, detect intrusions. There are different clustering approaches that can be used for anomaly detection. A summary of some of the suggested methods is below.

For grouping related data instances into clusters, clustering is used. While semi-supervised clustering has also been explored recently, clustering is mainly an unsupervised technique. While clustering and detection of anomalies appear to be fundamentally different from each other, several anomaly detection techniques based on clustering have been created. Techniques for anomaly detection based on clustering can be divided into three groups.

The first group of strategies focused on clustering relies on the following assumption:

Assumption: Regular data instances belong to a data cluster, while anomalies do not belong to any data cluster.

Techniques based on the above assumption apply a known algorithm based on clustering to the data set and declare as anomalous any data instance which does not belong to any cluster. Several clustering algorithms can be used, such as DBSCAN, ROCK, and SNN clustering, that do not force every data instance to belong to a cluster. FindOut is an extension of the WaveCluster algorithm that eliminates the clusters observed from the data and declares the residual instances as anomalies.

A downside of such methods is that they are not designed for anomalies to be found, because the primary objective of the clustering algorithm is to find clusters.

The second group of strategies focused on clustering relies on the following assumption:

Assumption: Usual instances of data lie close to their nearest centroid cluster, while exceptions are far away from their nearest centroid cluster.

Two measures are made up of techniques based on the above assumption. The data is clustered using a clustering algorithm in the first step. In the second stage, its distance to its closest cluster centroid is calculated as its anomaly score for each data case.

Notice that the techniques mentioned above would not be able to identify such anomalies if the anomalies in the data type clusters alone.

A third class of clustering-based techniques based on the following assumption has been proposed to solve this problem:

Assumption: Large and dense clusters belong to regular data instances, while anomalies either belong to small or sparse clusters.

Techniques based on the above assumption declare as anomalous instances belonging to clusters the size and/or density of which is below a threshold.

2.4.1 K-Means

K-Means clustering is a method of cluster analysis where k disjoint clusters are described on the basis of the function value of the objects to be clustered. Here, k is a parameter specified by the user.

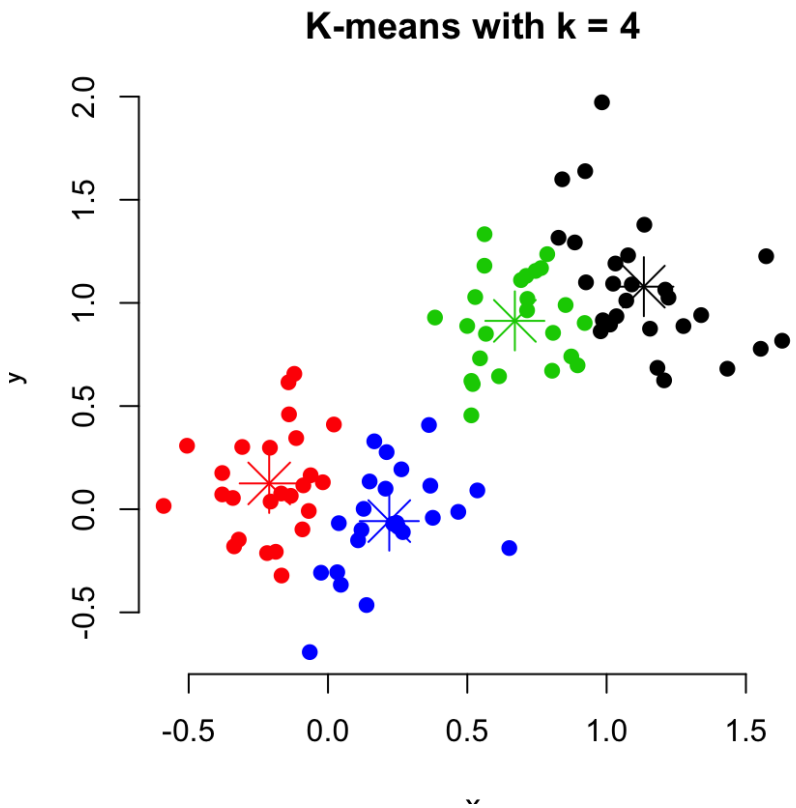


Figure 8

2.4.2 K-Medians

K-Medians is another K-Means-related clustering algorithm, except that we use the median group vector instead of recomputing the group centre points using the mean. This approach is less vulnerable to outliers (due to the use of the Median), but for larger datasets it is much slower as sorting is needed for each iteration when the Median vector is computed.

This algorithm is very similar to the algorithm for k-Means. It primarily differs in its representation of the various clusters. Each cluster is represented here by the most central object in the

Instead of the tacit means that does not belong to a cluster, the cluster. In the presence of noise and outliers, the k-medoids approach is more robust than the k-means algorithm since a medoid is less affected by outliers or other extreme values than a mean. This technique recognises network anomalies containing an unknown intrusion. It has been compared with numerous other clustering algorithms and has been discovered to deliver much better results than k-Means when it comes to accuracy.

2.4.3 EM Clustering

The technique of EM (expectation maximisation) is similar to the K-Means technique. The basic operation of the K-Means clustering algorithms is relatively simple: assign observations to those clusters, given a fixed number of k clusters, so that the means across clusters (for all variables) vary as much as possible from each other. This basic approach to clustering is expanded by the EM algorithm in two significant ways:

The EM clustering algorithm measures probabilities of cluster memberships based on one or more probability distributions rather than assigning examples to clusters to optimise the differences in means for continuous variables. The objective of the clustering algorithm then, provided the (final) clusters, is to maximise the overall likelihood or likelihood of the data.

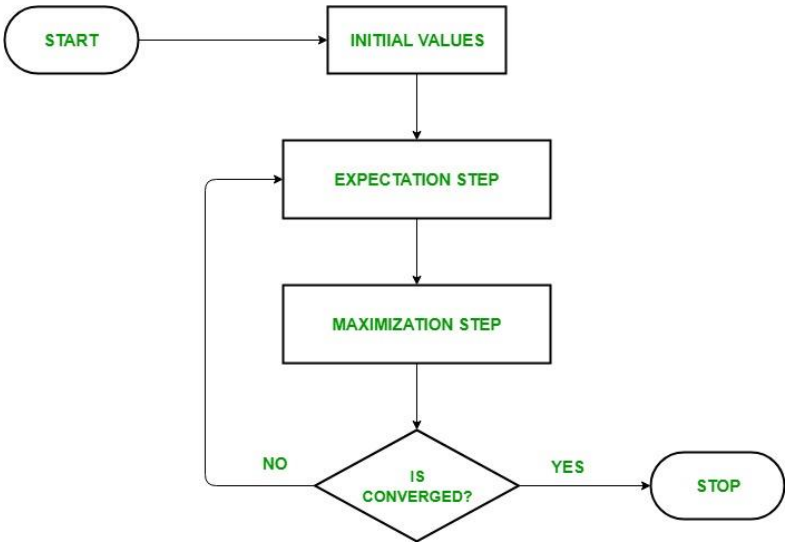


Figure 9

Advantages and Disadvantages of Clustering Based Techniques

The benefits of strategies focused on clustering are as follows:

i. Techniques based on clustering can work in an unsupervised mode.

ii. By simply plugging in a clustering algorithm that can handle the specific data type, such techniques can also be generalised to other complex data types.

iii. The testing process for clustering-based techniques is fast since a small constant is the number of clusters to which each test instance needs to be compared.

The drawbacks of strategies focused on clustering are as follows:

i. The efficiency of clustering-based techniques depends heavily on the efficacy of the clustering algorithm in capturing typical instances of the cluster structure.
ii. As a by-product of clustering, many techniques detect anomalies and are thus not designed for anomaly detection.
iii. Multiple clustering algorithms force some cluster to be allocated to each case. This could lead to anomalies being allocated to a large cluster, and techniques operating under the presumption that anomalies may not belong to any cluster are thus treated as usual cases.
iv. Only when the anomalies do not shape significant clusters among themselves are several clustering based techniques successful.
v. Computational complexity for data clustering is often a bottleneck, especially when using O(N2d) clustering algorithms.

2.5 Other Techniques for Anomaly Detection

A variety of methods of anomaly detection presented by different researchers have been presented. Some of these are listed below.

2.5.1 Graph based Anomaly detection

Subdue is an algorithm inside graphs for the identification of repeated patterns (substructures). The aim of the detection of anomalous substructures is to analyse an entire graph and report unusual substructures found within it. This sounds straightforward, but some subtleties are involved. For instance, merely searching for substructures that occur infrequently is not enough, because very large substructures are supposed to occur infrequently. (For example, it can not occur more than once if we consider the entire graph as a substructure.)

In essence, the Subdue system is a process for identifying patterns within graphs. The secret to the technique is that an anomaly should be called the "opposite" of a pattern — just as patterns always occur in a graph, anomalies occur infrequently. Large substructures (e.g., the

entire graph) will not be flagged as anomalous since they will be very large in size, and only if they do not occur very often will single-vertex substructures be considered anomalous.
It is necessary to note that this measure is biased towards the detection of very tiny substructures. This is because it is supposed to occur just a few times in larger substructures; the smaller the substructure, the less likely it is to be uncommon.

2.5.2 k-means and Sequential Minimal Optimization (SMO)

In the clustering process, two clusters were defined and generated by applying the K-means clustering algorithm. Each cluster's architecture is transferred to another as the algorithm iterates through the training data. Cluster updating allows the values of the centroids to be changed. This shift is a result of the elements of the new cluster. The clustering of the K-Means algorithm becomes complete when there are no changes to either cluster.

Sequential Minimal Optimization (SMO) was used to identify the dataset as normal or anomaly in the classification process supervised algorithm. Sequential Minimal Optimization (SMO) is a simple algorithm that can solve the SVM QP problem quickly without any additional storage of the matrix and without using steps to optimise numerical QP at all. The overall QP issue is broken down into QP sub-problems by SMO. The value of SMO resides in the fact that it is possible to solve two Lagrange multipliers analytically. Numerical QP optimization is therefore completely avoided.

A contrast was made between the proposed approach (K-mean + SMO) and the individual K-mean clustering algorithm and the classification of Sequential Minimal Optimization (SMO), and the results show that the approach outperforms others with a positive detection rate (94.48 percent) and decreases the false alarm rate (1.2 percent) and high precision (97.3695 percent)

2.5.3 Histogram-based Outlier Score (HBOS)

Using histograms to maintain a profile of the normal data is the simplest non-parametric statistical technique. These approaches are often referred to as frequency-dependent or dependent on counting. Histogram-based techniques are particularly common in the community of intrusion detection [Eskin 2000; Eskin et al. 2001; Denning 1987] and fraud

detection [Fawcett and Provost 1999], as certain profiles (user or software or system) that can be effectively captured using the histogram model regulate the conduct of the data.

Two steps consist of a simple histogram-based anomaly detection technique for univariate data. The first step involves the creation of a histogram based on the various values in the training data taken by that function. The technique tests in the second step if a test instance falls in any of the histogram bins. The test example is natural if it does, otherwise it is anomalous. A variation of the basic technique based on the histogram is to give each test instance an anomaly score based on the height (frequency) of the bin in which it falls.

For anomaly detection, the size of the bin used when constructing the histogram is important. Many standard test cases fall into empty or uncommon bins if the bins are small, resulting in a high false alarm rate. Many anomalous test cases will fall into frequent bins if the bins are big, resulting in a high false negative rate. Therefore, deciding an appropriate size of the bins to create the histogram, which maintains a low false alarm rate and low false negative rate, is a key challenge for histogram-based techniques.

2.5.4 SVM(One class and Two class)

Two components exist. The first one is the commitment of detectors. The detector is actively learning and self-evolving. The detector can measure an abnormality score for a newly collected data record. An alert will be activated if the score is below a threshold, likely with the kind of abnormality that can help a system administrator locate the anomaly. The second aspect is retraining the detector and selection of the working data collection. If some new data records are included in the working data collection, the detector needs to be retrained. The data set's size would expand rapidly from zero to something very high. All data records are initially standard. A small number of odd records can surface as time goes on. It is possible to mark certain irregular records according to their forms of anomaly.

Centered on 1-class and 2-class SVMs for adaptive failure detection, this Hybrid Anomaly Detection mechanism was used. It does not require a previous failure history, unlike other failure detection methods, and it can adapt itself by learning at runtime from observed failure events. It is therefore capable of discovering failures not seen in the past yet.

2.6 Relative Strengths and Weaknesses of Anomaly Detection Techniques

The unique strengths and limitations of each of the vast number of anomaly detection techniques discussed in previous sections. It is important to know which technique of anomaly detection is better suited for a given problem of anomaly detection. Given the complexity of the problem space, for any anomaly detection problem, it is not feasible to have such an understanding. For a few basic problem settings, we examine the relative strengths and limitations of various categories of techniques here.

Different kinds of techniques face distinct challenges with more complex data sets. When the number of dimensions is high, the nearest neighbour and clustering-based strategies suffer because the distance measures do not distinguish between normal and anomalous instances in a large number of dimensions. By mapping data to a lower dimensional projection, spectral techniques specifically answer the high dimensionality problem. Their efficiency, however, is highly dependent on the assumption that the usual instances and anomalies in the projected space are distinguishable. In such a situation, classification-based techniques can be a safer option. Classification-based approaches, however, require labels for both usual and anomalous situations, which are often not available, to be the most successful. Even if there are labels available for both usual and anomalous cases, the imbalance in the distribution of the two labels also makes it very difficult to learn a classifier. Semi-supervised nearest neighbour and clustering techniques can also be more efficient than classification-based techniques, which only use normal labels. Statistical approaches, while unsupervised, are only successful when the data dimensionality is low and statistical assumptions are held. Knowledge theoretical techniques require a test that is sufficiently sensitive to detect even a single anomaly's impact. Otherwise, such methods can only detect anomalies when there are a large number of anomalies.

Techniques based on nearest neighbour and clustering involve distance computation between a pair of data instances. Such techniques therefore conclude that the distance measure will distinguish well enough between the anomalies and usual instances. In cases where it is difficult to identify a good distance metric, a better alternative may be classification-based or statistical techniques.

A key aspect of an anomaly detection technique is its computational complexity, especially when the technique is applied to a real domain. Although there are costly training times for

classification-based, clustering-based and statistical techniques, testing is typically easy. This is also appropriate, as models can be trained in an offline fashion while real-time testing is required. Techniques such as nearest neighbour-based, data theoretical, and spectral techniques that do not have a training stage, on the other hand, have a costly testing stage that can be a limitation in a real world.

Typically, anomaly detection methods presume that data irregularities are uncommon as opposed to normal events. Anomalies are not necessarily rare, although this statement is usually valid. Anomalous (worm) traffic is actually more regular than usual traffic when dealing with worm detection on computer networks , for example. For such bulk anomaly detection, unsupervised techniques are not suited.

CHAPTER 3

CONCLUSION AND FUTURE SCOPE

Different data mining techniques are listed in this paper for the anomaly detection proposed in the past few years. This analysis would be useful for researchers to gain a basic understanding of different anomaly detection approaches. While much work has been done using independent algorithms, hybrid methods are commonly used as they have better results and resolve one approach's disadvantage over the other. New unknown attacks are witnessed every day and there is therefore a need for certain approaches that can detect the unknown activity stored, transferred or changed in the data set. Fusion or combination of already existing algorithms that have been proposed are listed in this research work.

We addressed various ways in which the issue of anomaly detection was formulated in the literature in this survey and tried to provide an overview of the enormous literature on different techniques. We also established a particular assumption regarding the notion of natural and anomalous data for each type of anomaly detection techniques. These assumptions may be used as guidance when applying a particular technique to a specific domain to determine the efficacy of the technique in that domain. Ideally, a detailed anomaly detection survey should allow a reader not only to understand the rationale behind the use of a specific technique for anomaly detection, but also to provide a comparative overview of different techniques. But the current analysis has been conducted in an unstructured manner, without relying on a single notion of anomalies, which makes it very difficult to provide a theoretical understanding of the issue of anomaly detection. A potential future task will be to unify into a mathematical or machine learning system the assumptions made by various techniques about normal and anomalous behaviour. Knorr and Ng [1997] include a small attempt in this direction, where the authors illustrate the relationship between distance-based and statistical anomalies for two-dimensional data sets.

In anomaly detection, there are several promising directions for further study. In many domains, contextual and collective anomaly detection techniques are beginning to find growing applicability and there is a great deal of potential for new techniques to be developed in this field. The need for distributed anomaly detection techniques has been inspired by the presence of data across various distributed locations. While such methods process information accessible at multiple sites, they also have to protect the information present at each site at the same time, requiring anomaly detection techniques to maintain privacy. Processing data as it arrives has become a requirement with the advent of sensor networks. Until detecting abnormalities, several techniques addressed in this survey involve the entire test data. Techniques that can function in an online manner have recently been suggested; such methods not only allocate an anomaly score to a test instance as it arrives, but also update the model incrementally. In complex structures, another upcoming field in which anomaly detection finds more and more applicability is. An example of such a system will be a multi-component aircraft system. In such systems, anomaly detection involves modelling the interaction between different components.

CHAPTER 4

REFERENCES

[1] Amanpreet Chauhan, Gaurav Mishra, Gulshan Kumar "Survey on Data Mining Techniques in Intrusion Detection" International Journal of Scientific & Engineering Research Volume 2, Issue 7, July-2011.

[2] Roshan Chitrakar, Huang Chuanhe "Anomaly based Intrusion Detection using Hybrid Learning Approach of combining k-Medoids Clustering and Naïve Bayes Classification" In Proceedings of 8th IEEE International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM); 2012.

[3] Varun Chandola, Arindam Banerjee, Vipin Kumar "Anomaly Detection: A Survey" ACM Computing Surveys (CSUR).

[4] Shikha Agrawal, Jitendra Agrawal "Survey on Anomaly Detection using Data Mining Techniques"

[5] Gerhard Munz, Sa Li, Georg Carle "Traffic Anomaly Detection Using K-Means Clustering"

[6] Chih-Fong Tsai , Yu-Feng Hsu , Chia-Ying Lin , Wei-Yang Lin "Intrusion detection by machine learning" Expert systems with applications, 2009 - Elsevier

[7] Santosh Nirmal, Maharashtra, India "Comparative Study between K-Means and K-Medoids Clustering Algorithms"

[8] Fu Song, Liu Jianguo and Pannu Husanbir "A Hybrid Anomaly Detection Framework in Cloud Computing Using One-Class and Two-Class Support Vector Machines"

[9] Syarif Iwan, Prugel-Bennet Adam, Wills Gary "Data Mining Approaches for Network Intrusion Detection: from Dimensionality Reduction to Misuse and Anomaly Detection"

[10] C. Cable, J. Cook Diane "Graph-Based Anomaly Detection"

[11] Gadal Saad Mohamed Ali Mohamed, Mokhtar Rania A. "Anomaly Detection Approach using Hybrid Algorithm of Data Mining Technique"

[12] Kruegel Christopher, Vigna Giovanni "Anomaly Detection of Web-based Attacks"

[13] Goldstein Markus, Uchida Seiichi "A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data"

[14] Stefan Holban "A Genetic Algorithm for Classification "

[15] Syarif Iwan, Prugel-Bennet Adam, Wills Gary "Unsupervised Clustering Approach for Network Anomaly Detection"

[16] Bridges Susan M., Naughm Rayford B. “Intrusion Detection Via Fuzzy Data Mining”

[17] Goldstein Markus, Dengel Andreas “Histogram-based Outlier Score (HBOS): A fast Unsupervised Anomaly Detection Algorithm”

[18] Platt John C. “Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines”

[19] Siaterlis, C. and Maglaris, B. 2004. Towards multisensor data fusion for dos detection. In Proceedings of the 2004 ACM symposium on Applied computing. ACM Press, 439–446.

[20] Ramadas, M., Ostermann, S., and Tjaden, B. C. 2003. Detecting anomalous network traffic with self-organizing maps. In Proceedings of Recent Advances in Intrusion Detection.

[21] Chandola, V., Eilertson, E., Ertoz, L., Simon, G., and Kumar, V. 2006. Data mining for cyber security. In Data Warehousing and Data Mining Techniques for Computer Security, A. Singhal, Ed. Springer.

[22] Bolton, R. and Hand, D. 1999. Unsupervised profiling methods for fraud detection. In Credit Scoring and Credit Control VII.

[23] Breunig, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J. 2000. Lof: identifying density-based local outliers. In Proceedings of 2000 ACM SIGMOD International Conference on Management of Data. ACM Press, 93–104.

[24] Fawcett, T. and Provost, F. 1999. Activity monitoring: noticing interesting changes in behavior. In Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM Press, 53–62.

[25] Phua, C., Alahakoon, D., and Lee, V. 2004. Minority report in fraud detection: classification of skewed data. SIGKDD Explorer Newsletter 6, 1, 50–59.

[26] Taniguchi, M., Haft, M., Hollmn, J., and Tresp, V. 1998. Fraud detection in communications networks using neural and probabilistic methods. In Proceedings of IEEE International Conference in Acoustics, Speech and Signal Processing. Vol. 2. IEEE Computer Society, 1241–1244.

[27] Lin, J., Keogh, E., Fu, A., and Herle, H. V. 2005. Approximations to magic: Finding unusual medical time series. In Proceedings of the 18th IEEE Symposium on Computer-Based Medical Systems. IEEE Computer Society, Washington, DC, USA, 329–334.

[28] Aggarwal, C. 2005. On abnormality detection in spuriously populated data streams. In Proceedings of 5th SIAM Data Mining. 80–91.

[29] Diaz, I. and Hollmen, J. 2002. Residual generation and visualization for understanding novel process conditions. In Proceedings of IEEE International Joint Conference on Neural Networks. IEEE, Honolulu, HI, 2070–2075.

[30] Yairi, T., Kato, Y., and Hori, K. 2001. Fault detection by mining association rules from housekeeping data. In In Proceedings of International Symposium on Artificial Intelligence, Robotics and Automation in Space.

[31] Srivastava, A. 2006. Enabling the discovery of recurring anomalies in aerospace problem reports using high-dimensional clustering techniques. Aerospace Conference, 2006 IEEE, 17–34.

[32] Manevitz, L. M. and Yousef, M. 2000. Learning from positive data for document classification using neural networks. In Proceedings of Second Bar-Ilan Workshop on Knowledge Discovery and Learning. Jerusalem.