


# An Efficient Mixed Attribute Outlier Detection Method for Identifying Network Intrusions

J. Rene Beulah, Saveetha School of Engineering, India

 <https://orcid.org/0000-0001-9246-1522>

D. Shalini Punithavathani, Government College of Engineering, India

## ABSTRACT

Intrusion detection systems (IDS) play a vital role in protecting information systems from intruders. Anomaly-based IDS has established its effectiveness in identifying new and unseen attacks. It learns the normal usage pattern of a network and any event that significantly deviates from the normal behavior is signaled as an intrusion. The crucial challenge in anomaly-based IDS is to reduce false alarm rate. In this article, a clustering-based outlier detection (CBOD) approach is proposed for classifying normal and intrusive patterns. The proposed scheme operates in three modules: an improved hybrid feature selection phase that extracts the most relevant features, a training phase that learns the normal pattern in the training data by forming clusters, and a testing phase that identifies outliers in the testing data. The proposed method is applied for NSL-KDD benchmark dataset and the experimental results yielded a 97.84% detection rate (DR), a 1.88% false alarm rate (FAR), and a 97.96% classification accuracy (ACC). This proposal appears to be promising in terms of DR, FAR and ACC.

## KEYWORDS

Clustering, Feature Selection, NSL-KDD, Outlier, Profile

## 1. INTRODUCTION

Intrusion detection systems (IDS) addresses problems that are not solved by firewall techniques. IDS is capable of recognizing those attacks which firewalls are not able to prevent. (Pradhan et al., 2016). IDS can detect malicious activities performed by external or internal attackers (Korba et al., 2017). These attacks attempt to disrupt legitimate user's access to services (Nagesh et al., 2017). There is a growing need for efficient methods to detect outliers or anomalies in network traffic data. Network traffic data is massive and highly dimensional, and it is challenging to extract relevant information to identify attacks. Anomaly based IDS is of great interest in the research community for many years and is based on the assumption that the behavior of intruders is different from that of a legitimate user. Anomaly based IDS can easily attain very high detection rate using a strict definition of normal activities, but at the cost of unacceptably higher false alarm rate. It is in fact a challenging task to improve DR beyond a certain limit while keeping FAR at a reduced level.

Outlier detection is a data mining concept that finds immense applications in varied fields. Outliers are data that are notably different from the rest of the data. Originally, outlier detection was used as a preprocessing step for removing noise and extreme values. But nowadays, outlier detection

DOI: 10.4018/IJISP.2020070107

has become a field of interest for applications involving detection of fraudulent activities as it can be used to isolate suspicious patterns (Settanni & Filzmoser, 2018; Settanni & Filzmoser, 2018; Domingues et al., 2018; Rousseeuw et al., 2019). Outlier detection has been used for centuries to detect and remove anomalous data points (Hodge, 2014). Anomaly-based IDS is an application area where outlier analysis plays a vital role because intrusions are rare events compared to normal events and these rare events can be treated as outliers. “The anomaly detection problem is similar to the problem of finding outliers, specifically in network intrusion detection” (Gogoi, Borah, Bhattacharyya & Kalita, 2011). A hacker inside a network with an evil intent can be pointed out obviously by an outlier (Ganapathy, Jaisankar, Yokesh & Kannan, 2011).

There are many methods employed in the literature for outlier detection like statistical-based, distance-based, density-based, clustering-based and frequent-pattern-based. A detailed survey of such outlier detection methods applied for detecting network intrusions is given in (Beulah & Punithavathani, 2015). All the outlier detection methods create a model for the normal pattern in the data and then outliers are identified by finding the deviations from the learned model.

Clustering can be regarded as a complimentary problem to outlier detection (Aggarwal, 2015). Clustering aims at finding data points having similar properties whereas outlier detection looks for data points that are different from others. In most of the clustering algorithms outliers are obtained as side-products. Clustering is much suitable for the problem of network intrusion detection and is one of the most effective ways to decide whether a connection is legitimate or malicious (Hassani & Seidl, 2011). Clustering algorithms can be carefully designed to detect outliers or anomalies in network traffic data.

From the viewpoint of a clustering algorithm, outliers are objects not located in clusters of a dataset (Bakar, Mohemad, Ahmad & Deris, 2006). Clustering-based outlier detection approaches use some clustering algorithm to divide the data set into clusters. A data point that does not strongly belong to any cluster is an outlier (single point outlier). Some very small clusters can also be outliers (cluster-based outliers). Even though there are many approaches for finding outliers, clustering is the most common and natural way. It is a highly efficient approach in terms of space and time complexity.

The main contribution of this work is a novel clustering-based approach for identifying outliers in any kind of data. It forms clusters in a very meaningful way and classifies data points that do not belong to any cluster and small clusters as outliers. The method is employed for identifying intrusions in network traffic connections and the results are presented in detail. The main objective behind this work is to reduce FAR as much as possible while maintaining high DR. This is achieved using an efficient hybrid feature selection method that has well extracted the most relevant features and a clustering-based outlier detection approach that distinguishes normal and intrusive activities more accurately.

## **2. RELATED WORK**

Anomaly-based IDS is the prime focus of research in the field of intrusion detection since it was originally proposed by Denning, 1987. Many proposals for network anomaly detection can be found in the literature for the past few decades. An overview of research in anomaly detection schemes, its challenges, assessment issues, classification of anomaly detection schemes and comparison of supervised and unsupervised approaches can be found in (Lazarevic et al., 2003; Chandola et al., 2009; Gracia-Teodoro et al., 2009; Zhang et al., 2009; Ghorbani et al., 2009; Tavallae et al., 2010; Gogoi et al., 2010; Bhuyan et al., 2011; and Bhuyan et al. 2014). A common problem faced by anomaly-based IDS is the phenomena of increased false alarms.

Data mining is a rapidly developing research area that helps in automatically bringing out useful information hidden in huge data stores. Lee & Stolfo were the first to apply data mining techniques for detecting intrusions (Lee & Stolfo, 1998). Since then, data mining methods have found a prominent place in the literature and have demonstrated high accuracy. Data mining is one of the most known

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

[www.igi-global.com/article/an-efficient-mixed-attribute-outlier-detection-method-for-identifying-network-intrusions/256571?camid=4v1](http://www.igi-global.com/article/an-efficient-mixed-attribute-outlier-detection-method-for-identifying-network-intrusions/256571?camid=4v1)

This title is available in InfoSci-Journals, InfoSci-Journal Disciplines Computer Science, Security, and Information Technology, InfoSci-Journal Disciplines Engineering, Natural, and Physical Science, InfoSci-Computer Science and IT Knowledge Solutions – Journals, InfoSci-Knowledge Discovery, Information Management, and Storage eJournal Collection. Recommend this product to your librarian:

[www.igi-global.com/e-resources/library-recommendation/?id=2](http://www.igi-global.com/e-resources/library-recommendation/?id=2)

## Related Content

---

Personalized Key Drivers for Individual Responses in Regression Modeling  
Stan Lipovetsky (2020). *International Journal of Risk and Contingency Management* (pp. 15-30).

[www.igi-global.com/article/personalized-key-drivers-for-individual-responses-in-regression-modeling/252179?camid=4v1a](http://www.igi-global.com/article/personalized-key-drivers-for-individual-responses-in-regression-modeling/252179?camid=4v1a)

End-to-End (E2E) Security Approach in WiMAX: A Security Technical Overview for Corporate Multimedia Applications

Sasan Adibi, Gordon B. Agnew and Tom Tofigh (2008). *Handbook of Research on Wireless Security* (pp. 747-758).

[www.igi-global.com/chapter/end-end-e2e-security-approach/22082?camid=4v1a](http://www.igi-global.com/chapter/end-end-e2e-security-approach/22082?camid=4v1a)

## Copyright Protection in Virtual Communities through Digital Watermarking

Huayin Si and Chang-Tsun Li (2008). *Information Security and Ethics: Concepts, Methodologies, Tools, and Applications* (pp. 3788-3793).

[www.igi-global.com/chapter/copyright-protection-virtual-communities-through/23327?camid=4v1a](http://www.igi-global.com/chapter/copyright-protection-virtual-communities-through/23327?camid=4v1a)

## Would Be Pirates: Webcasters, Intellectual Property, and Ethics

Melanie J. Mortensen (2008). *Information Security and Ethics: Concepts, Methodologies, Tools, and Applications* (pp. 380-401).

[www.igi-global.com/chapter/would-pirates-webcasters-intellectual-property/23100?camid=4v1a](http://www.igi-global.com/chapter/would-pirates-webcasters-intellectual-property/23100?camid=4v1a)