Chapter 8
Anomaly Detection in Wireless Networks: An Introduction to Multi-Cluster Technique

Yirui Hu
Rutgers University, USA

ABSTRACT

This chapter is an introduction to multi-cluster based anomaly detection analysis. Various anomalies present different behaviors in wireless networks. Not all anomalies are known to networks. Unsupervised algorithms are desirable to automatically characterize the nature of traffic behavior and detect anomalies from normal behaviors. Essentially all anomaly detection systems first learn a model of the normal patterns in training data set, and then determine the anomaly score of a given testing data point based on the deviations from the learned patterns. The initial step of learning a good model is the most crucial part in anomaly detection. Multi-cluster based analysis are valuable because they can obtain the insights of human behaviors and learn similar patterns in temporal traffic data. The anomaly threshold can be determined by quantitative analysis based on the trained model. A novel quantitative “Donut” algorithm of anomaly detection on the basis of model log-likelihood is proposed in this chapter.

DOI: 10.4018/978-1-5225-1750-4.ch008
INTRODUCTION

An anomaly in wireless networks represents an unusual traffic pattern that deviates from the normal (or usual) network behavior. Anomalies are also referred to as abnormalities, deviants, or outliers in data mining and statistics field. Network anomalies can be root caused by a variety of issues such as introduction of new features, network intrusions or disaster events. In many cases, intrusion for example, the outliers can only be discovered as a sequence of multiple data points, rather than as an individual data point.

The general steps in developing an anomaly detection system are: first define a model of the normal patterns using training data; then compute the deviation score of testing data points given normal patterns (Aggarwal, 2013). Training a model for co-occurrence data includes both manually and automatic techniques. In past few decades, the network alarm thresholds can be set manually and tuned by subject matter experts (SME) for each generating entity. In nowadays network traffic, however, it is almost impossible for SMEs to manually detect anomalies from such voluminous amount of high-dimensional data. Effective anomaly detection systems based on machine learning algorithms are hence desirable to automatically extract useful information in terms of abnormal characteristics of the systems and entities from such noisy, high-dimensional data and provide useful application-specific insights.

It is therefore crucial to learn a good model based on training data. The optimal choice of a model is often data set specific, which requires a good understanding of the particular domain before choosing the model (Aggarwal, 2013). Clustering analysis can be a promising candidate in learning similar patterns in temporal traffic data.

BACKGROUND

The initial step of learning a good model is the most crucial part in anomaly detection. In general, an incorrect choice of model can lead to poor anomaly detection results. For example, a linear model may not work well if the underlying pattern is generated from multiple clusters. In such cases, the testing data can be mistakenly detected as anomaly because the poor fit to the learned linear model, which lead to high false alarm rates. Effective anomaly detection systems based on machine learning algorithms are hence desirable to automatically extract useful information in terms of abnormal characteristics of the systems and entities from such noisy, high-dimensional data and provide useful application-specific insights.

Machine learning algorithms include supervised learning and unsupervised learning (Jain, Murty, & Flynn, 1999; Theodoridis & Koutroumbas, 2006). Most anomaly detection systems employ supervised algorithms based on training data,
Related Content

Performance Analysis of On-Chip Communication Structures under Device Variability
[www.igi-global.com/article/performance-analysis-chip-communication-structures/47541?camid=4v1a](www.igi-global.com/article/performance-analysis-chip-communication-structures/47541?camid=4v1a)

The Adoption of Broadband Internet in Australia and Canada
[www.igi-global.com/chapter/adoption-broadband-internet-australia-canada/20476?camid=4v1a](www.igi-global.com/chapter/adoption-broadband-internet-australia-canada/20476?camid=4v1a)
Location Management in PCS Networks Using Base Areas (BAs) and 2 Level Paging (2LP) Schemes
[www.igi-global.com/article/location-management-pcs-networks-using/55101?camid=4v1a](www.igi-global.com/article/location-management-pcs-networks-using/55101?camid=4v1a)

Strategic Moves Related to Broadband Diffusion on the French VoD and TV Market
[www.igi-global.com/chapter/strategic-moves-related-broadband-diffusion/48308?camid=4v1a](www.igi-global.com/chapter/strategic-moves-related-broadband-diffusion/48308?camid=4v1a)