

# DETECÇÃO DE ANOMALIAS EM REDES: TAXONOMIA DE TÉCNICAS E UM ESTUDO DE CASO

ANOMALY DETECTION NETWORKS: TECHNIQUES TAXONOMY AND A CASE  
STUDY

Bruno Prece da Silva<sup>1</sup>

Me. Mario Henrique Akihiko da Costa Adaniya.<sup>2</sup>

## Resumo

A detecção de anomalias em redes é uma área de pesquisa muito dinâmica, que envolve diversas técnicas para detectar anomalias conhecidas e desconhecidas. Diversos sistemas e algoritmos são propostos e há muitas discussões referentes à eficácia de cada método na literatura. Este artigo apresenta alguns aspectos da detecção de anomalias, além disso, apresenta um estudo de caso utilizando um algoritmo simples e eficaz para o agrupamento de dados.

**Palavras-chave:** Detecção de Anomalias; Clustering; Segurança da Informação.

## Abstract

The detection of anomalies in networks is an area of very dynamic research that involves various techniques to detect known and unknown anomalies. Several systems and algorithms are proposed and there are many discussions concerning the effectiveness of each method in the literature. This paper presents some aspects of anomaly detection, in addition, presents a case study using a simple and effective algorithm for grouping data..

**Keywords:** Anomaly Detection; Clustering; Information Security.

## INTRODUÇÃO

A constante expansão na utilização da Internet tornou a rede um serviço essencial para a interconexão global e para as empresas. A cada dia há um crescimento na demanda por novos serviços, que necessitam de políticas de segurança que mantenham a integridade e a privacidade dos dados. Isso torna os sistemas mais complexos e os dados cada vez mais heterogêneos, impossibilitando que a gerência da rede seja realizada por um operador humano. Devido à estas dificuldades, surgem comportamentos anômalos no tráfego.

---

<sup>1</sup>Centro Universitário Filadélfia de Londrina - UniFil

<sup>2</sup>Centro Universitário Filadélfia de Londrina - UniFil

## DETECÇÃO DE ANOMALIAS

No tráfego de rede as anomalias representam ações que se diferem do comportamento normal que já foi analisado. Mesmo quando uma anomalia não impacta profundamente a rede, ela atinge a qualidade dos serviços que são entregues aos usuários finais (LAKHINA; CROVELLA; DIOT, 2004). O comportamento anômalo pode ser gerado por eventos de origem maliciosa, como, por exemplo, ataques de negação de serviço, varredura de portas e ataques de penetração. Além de eventos não maliciosos, como, aumentos repentinos no volume do tráfego, tempestades de *broadcast* e congestionamentos.

As principais vertentes de pesquisas dividem as técnicas de detecção de anomalias em duas categorias, a detecção baseada em assinaturas de anomalias que se destaca em relação à detecção de anomalias já conhecidas, e detecção baseada na caracterização do comportamento normal da rede que possui um melhor desempenho na detecção de anomalias desconhecidas. Os sistemas de detecção são classificados de acordo com as assinaturas utilizadas na detecção. Sistemas supervisionados utilizam as duas classes de assinaturas, diferente dos sistemas semi-supervisionados fazem o uso apenas da assinatura do perfil normal do tráfego. E os sistemas não supervisionados que não utilizam assinaturas, sendo mais fáceis de serem aplicados na rede, porém, com maior dificuldade de implementação (CHANDOLA; BANERJEE; KUMAR, 2009).

Vários métodos e algoritmos são utilizados nas pesquisas e nos sistemas para detecção de anomalias. Dentre estas técnicas, a clusterização de dados é muito explorada e possui bons resultados. A clusterização consiste no particionamento e agrupamento de um conjunto de dados, de tal forma que os objetos que compartilhem características comuns e que se diferem de outros cluster sejam alocados em um mesmo grupo, assim, as técnicas de clustering precisam encontrar automaticamente a assinatura dos dados, ou seja, as características mais relevantes de cada grupo (REHMAN; REHMAN; KHAN, 2009).

## METODOLOGIA EXPERIMENTAL

Para a elaboração deste trabalho foi utilizado o K-means, um algoritmo simples e muito poderoso para clusterização de dados. A técnica divide um conjunto de  $n$

elementos em um número  $K$  de clusters, de um modo que a similaridade entre os elementos de um mesmo grupo seja alta e a similaridade entre os grupos seja baixa.

O algoritmo consiste primeiramente na escolha de  $K$  objetos centrais, onde  $K$  é a quantidade de clusters definidos por parâmetro. Assim, cada ponto de dado é atribuído ao seu centroide (objeto central) mais próximo, formando os grupos. Os centroides de cada cluster são atualizados de acordo com a média de similaridade dos objetos inseridos nos grupos, atualizando a média dos clusters a cada iteração. A atribuição de objetos nos clusters e a atualização das médias é realizada até que nenhum objeto seja inserido no cluster, e até que todos os centroides permaneçam os mesmos. Desta forma, o algoritmo realiza a clusterização de uma base de dados, agrupando os dados referentes ao comportamento normal do tráfego em um grupo e os dados referentes às anomalias serão agrupados em outros clusters de acordo com a característica de cada anomalia.

## REFERÊNCIAS

CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, ACM, v. 41, n. 3, p. 15, 2009. 2

LAKHINA, A.; CROVELLA, M.; DIOT, C. Diagnosing network-wide traffic anomalies. In: *Proceedings of the 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*. New York, NY, USA: ACM, 2004. (SIGCOMM '04), p. 219–230. ISBN 1-58113-862-8. Disponível em: <<http://doi.acm.org/10.1145/1015467.1015492>>. 2

REHMAN, Z.; REHMAN, S. A.; KHAN, L. Survey reports on four selected research papers on data mining based intrusion detection system. *University of the West Indies at Mona*, 2009. 2