Jan 2014, HAPPIEST MINDS TECHNOLOGIES

User Behavior based Anomaly Detection for Cyber Network Security

By

Sandip K Pal
Manish Anand

Happiest Minds, Analytics Practice

SHARING. MINDFUL. INTEGRITY. LEARNING. EXCELLENCE. SOCIAL RESPONSIBILITY.

Copyright Information

This document is an exclusive property of Happiest Minds Technologies Pvt. Ltd.It is intended for limited circulation.

Abstract

Fusion of data from multiple sources is generating new information from existing data. Now users can access any information from inside or outside of the organization very easily. It helps to increase the user productivity and knowledge shared within the organization. But this leads to a new area of network security threat, "Inside Threat". Now users can share critical information of organization to outside the organization if he/she has access to the information. The current network security tool cannot prevent the new threat. In this paper, we address this issue by "*Building real time anomaly detection system based on users' current behavior and previous behavior*".

Introduction

Securing data in a networked environment has been a major concern for Network Administrator and Security Personnel who are responsible for network security, as the networks are under threat byintruders. Generaing new information from existing data becomes more critical while integrating data from multiple sources. The intrusion may get access to value information and they might share/steal the information, causing damage to the network and applications running on the data base. Detection of anomaly is very essential as it is impossible to secure all the information in the network. To secure the information, there are many methods and techniques in place:

Authentication: Technology that requires user authentication before using the system. There are different levels of authentication like authenticating the user with user id and password (one factor authentication), security token (two factor authentication) or finger print (three factor authentication).

Packet Inspection: Examines the data part (and possibly also the header) of a packet as it passes an inspection point, searching for protocol non-compliance, viruses, spam, intrusions or defined criteria to decide whether the packet may pass or if it needs to be routed to a different destination or for the purpose of collecting statistical information.

Signature Detection: It monitors packets in the network and compares them with pre-configured and pre-determined attack patterns.

All three methods can detect anomaly in the network but they have low detection rate and high false alarm rate. They cannot detect new anomaly. Behavior based anomaly detection helps solve this problem.

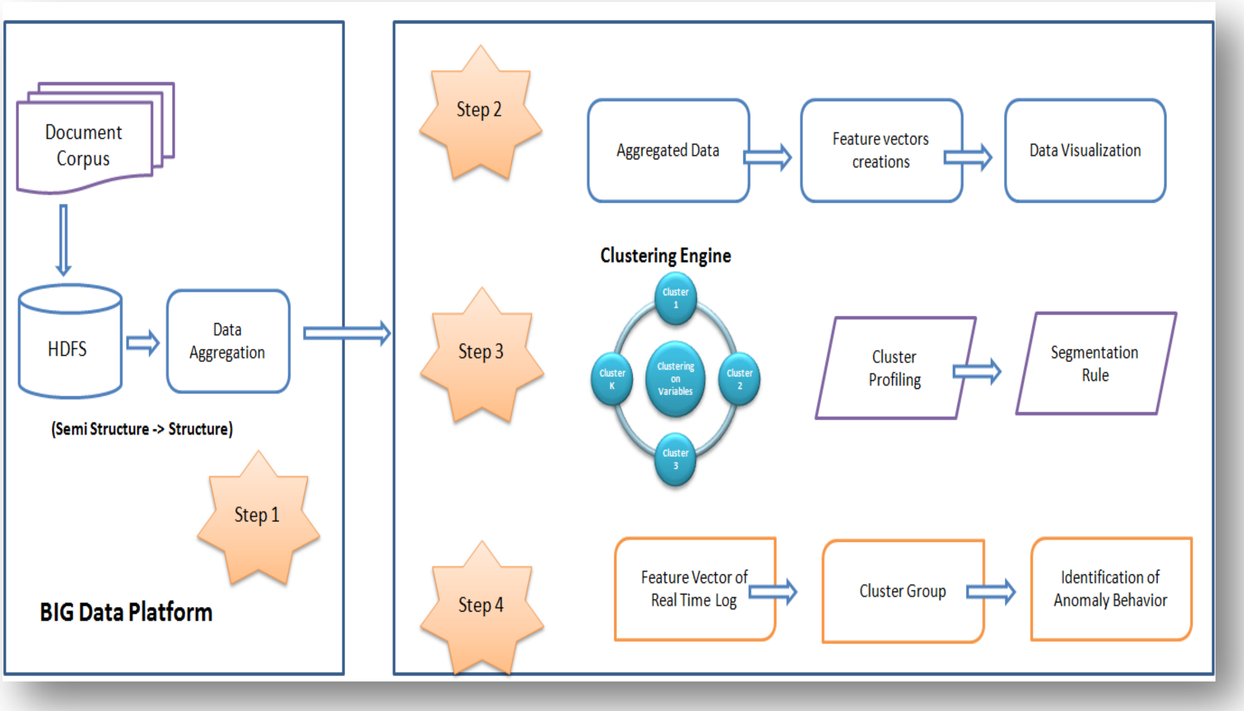
Network Behavior Anomaly Detection (NBAD) is a way to enhance the security of proprietary network by monitoring traffic and noting the unusual pattern or departure from normal behavior. It offers security, in addition to that provided by traditional anti-threat applications such as firewalls, antivirus software and spyware-detection software. NBAD Solution aggregates data from multiple sources and establishes the benchmark for normal behavior. After establishing a benchmark for normal traffic, it passively monitors network activity and flags unknown, new or unusual patterns that might indicate the presence of a threat. For example: when normal users access the network, they do it at human scale with a mouse and keyboard. If a hacker is accessing the network, it is very fast, with high-volume click rates and looking for different site/files and his/her login pattern is different. "Nothing about it looks like a normal user."

NBAD is also good at spotting new anomaly and zero day exploits. It helps a network administrator minimize the time and labor involved in locating and resolving problems.

Behavior Based Anomaly Detection Solution Framework:

System log data is massive in volume and semi–structured. It is very difficult and not cost effective to store data in normal data or processes using statistical tools. Store, process and aggregate data in Hadoop platform using map & reduce algorithm and use the aggregate data in statistical tools for data analysis and model building.

Proposed Solution Framework (Figure 1)



Approach:

Step1:

- Process the massive system log data in Hadoop platform
- Convert semi structured data to structured data
- Aggregate data using map & reduce algorithm

Step 2:

- Clean the data and create featured vector for each user
- Identify the outlier and do the missing value treatment
- Generate the insight for the data
- Identify the behavioral attribute for cluster

Step 3:

- Apply different cluster algorithms on massive data
- Select best methods and number of cluster users
- Cluster the users based on different behavioral attributes
- Cluster, profiling & identification of outlier cluster
- Rule generation for real time segmentation

Step 4:

- Aggregation of real time log and create feature vector for each user
- Real time user segmentation based on current user behavior and segmentation rules
- Identify the anomaly behavior user who belongs to outlier cluster

Case Study:

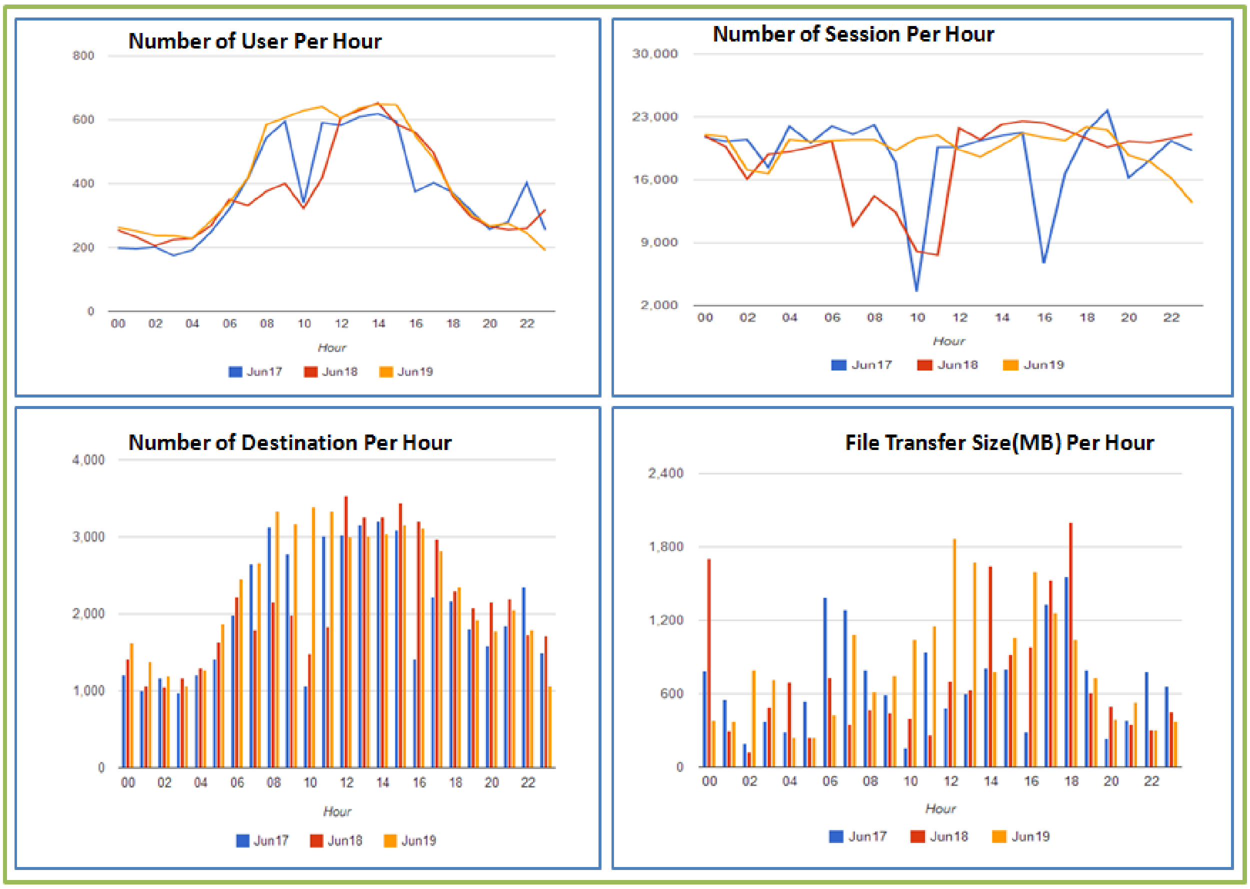
Executive Summary

There are thousands of users in an enterprise who connect systems everyday and generate massive system log data. They can access any information if they have authentication for that and can share the information with the outside world. Detection of anomaly from system log based on current rules is no more effective and it might miss some critical cases. The enterprise wanted to improve the anomaly detection rate and minimize the false alert rate.

The objective of this study was to detect anomaly from massive system log data based on user behavioral attributes like number of destinations, number of sessions, number of applications started by users, file transfer size, duration of sessions etc.

System log data used for this study contains all the action taken by users when connections were started with internal and/or external systems. Important fields of system log were Start time of the session, Source IP, Destination connecting location, firewall rule apply on the connection, application start by connection, file transfer size, session ID of connection, protocol applied on connection, action taken by firewall, packet transfer from source to destination, packet transfer from destination to source.

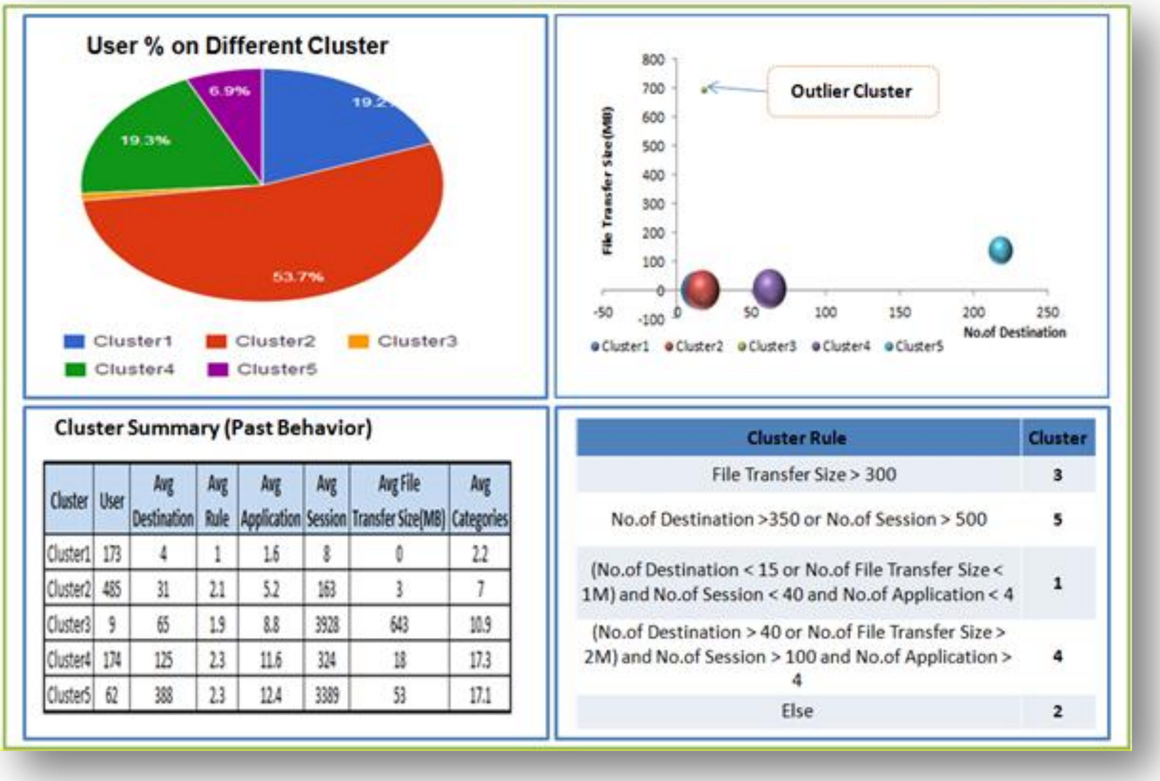
Data Description Output (Figure 2)



After data analysis we identified that:

- There were 903 users common across the days
- Most of them logged in between 8 AM - 2 PM
- There was a drastic drop in the number of users at 10AM on Jun17
- Maximum file transfers occurred between 4 PM and 6 PM

Cluster Output (Figure 3)



Cluster Summary (Past Behavior)

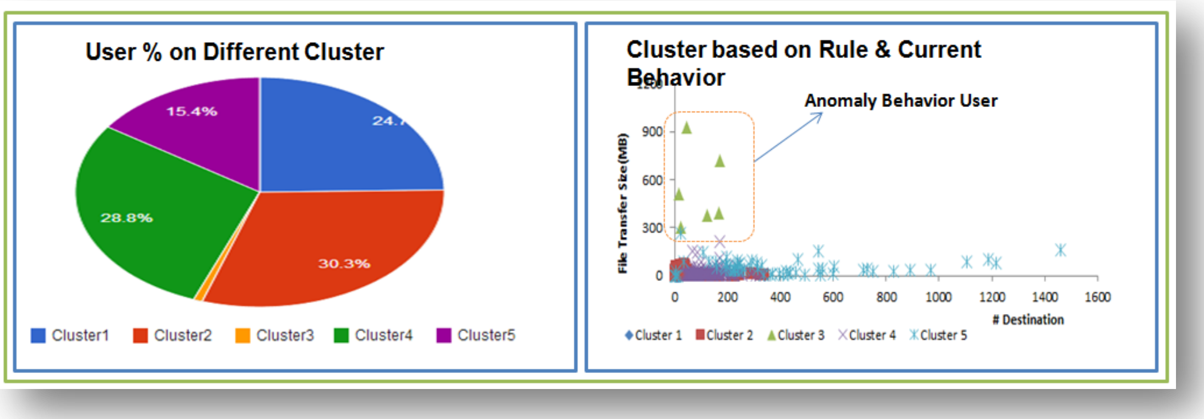
Cluster	User	Avg Destination	Avg Rule	Avg Application	Avg Session	Avg File Transfer Size(MB)	Avg Categories
Cluster1	173	4	1	1.6	8	0	2.2
Cluster2	485	31	2.1	5.2	163	3	7
Cluster3	9	65	1.9	8.8	3928	643	10.9
Cluster4	174	125	2.3	11.6	324	18	17.3
Cluster5	62	388	2.3	12.4	3389	53	17.1

Cluster Rule	Cluster
File Transfer Size > 300	3
No.of Destination >350 or No.of Session > 500	5
(No.of Destination < 15 or No.of File Transfer Size < 1M) and No.of Session < 40 and No.of Application < 4	1
(No.of Destination > 40 or No.of File Transfer Size > 2M) and No.of Session > 100 and No.of Application > 4	4
Else	2

Data Analysis revealed that were five segments based on the number of destination IPs, number of firewall rules applied on the users, number of applications started by users, number of sessions started by users, number of events completed by users, total file transferred size, number of categories.

Cluster 3 with high file transfer size and Cluster 5 with high number of destinations were statistically different from other clusters. Cluster 3 was the outlier cluster.

Anomaly Detection Output (Figure 4)



There were nine with anomaly behavior and their average file transfer size was more than 645 MB.

Conclusion:

Behavior based anomaly detection solution significantly increases the anomaly detection rate and minimizes the false alert rate. It also minimizes the time and labor involved in identification and resolving threats. The next step of this analysis is to build the prediction model to forecast threats with severity. The solution framework shown in the Figure 1 can be customized with specific requirements of each client. It can be implemented with any other platform without modifying the present processes and infrastructures significantly.

References:

1. "Using Strategy Objectives for Network Security Analysis" by Elie Burstein and John C. Mitchell
2. "Detecting APT Activity with Network Traffic Analysis" by Nart Villeneuve and James Bennett
3. "Design of intrusion detection system based of artificial neural network application of rough set" by Dilip Kumar Barman & Dr.Guruprasad Khataniar