

**CASE STUDY - NETWORKING NETWORK INTRUSION DETECTION**



## Website: [www.analytixlabs.co.in](http://www.analytixlabs.co.in/) Email: [info@analytixlabsl.co.in](mailto:info@analytixlabsl.co.in)

**Disclaimer:** This material is protected undercopyright act AnalytixLabs©, 2011-2018. Unauthorized use and/ or duplication of this materialorany part of this material including data, in any form without explicit and written permission from AnalytixLabs is strictly prohibited. Any violation of this copyright will attract legal actions.

AnalytixLabs, Website: www.a na l yti xl a bs .co.i n Email: i nfo@a na l yti xl a bs .co.i n phone: +91-88021-73069



**BUSINESS CONTEXT:**

With the enormous growth of computer networks usage and the huge increase in the number of applications running on top of it, network security is becoming increasingly more important. All the computer systems suffer from security vulnerabilities which are both technically difficult and economically costly to be solved by the manufacturers. Therefore, the role of Intrusion Detection Systems (IDSs), as special-purpose devices to detect anomalies and attacks in the network, is becoming more important.

The research in the intrusion detection field has been mostly focused on anomaly-based and misusebased detection techniques for a long time. While misuse-based detection is generally favored in commercial products due to its predictability and high accuracy, in academic research anomaly detection is typically conceived as a more powerful method due to its theoretical potential for addressing novel attacks.

Conducting a thorough analysis of the recent research trend in anomaly detection, one will encounter several machine learning methods reported to have a very high detection rate of 98% while keeping the false alarm rate at 1%. However, when we look at the state of the art IDS solutions and commercial tools, there is no evidence of using anomaly detection approaches, and practitioners still think that it is an immature technology. To find the reason of this contrast, lots of research was done done in anomaly detection and considered various aspects such as learning and detection approaches, training data sets, testing data sets, and evaluation methods.

**BUSINESS PROBLEM:**

Your task to build network intrusion detection system to detect anamolies and attacks in the network. There are two problems.

1. Binomial Classification: Activity is normal or attack
2. Multinomial classification: Activity is normal or DOS or PROBE or R2L or U2R

Please note that, currently the dependent variable (target variable) is not definied explicitly. However, you can use attack variable to define the target variable as required.

**DATA AVAILABILITY:**

This data is KDDCUP’99 data set, which is widely used as one of the few publicly available data sets

for network-based anomaly detection systems.

For more about data: <http://www.unb.ca/cic/datasets/nsl.html>

**LIST OF COLUMNS FOR THE DATA SET:**

["duration","protocol\_type","service","flag","src\_bytes","dst\_bytes","land", "wrong\_fragment","urgent","hot","num\_failed\_logins","logged\_in", "num\_compromised","root\_shell","su\_attempted","num\_root","num\_file\_creations", "num\_shells","num\_access\_files","num\_outbound\_cmds","is\_host\_login", "is\_guest\_login","count","srv\_count","serror\_rate", "srv\_serror\_rate", "rerror\_rate","srv\_rerror\_rate","same\_srv\_rate", "diff\_srv\_rate",

AnalytixLabs, Website: www.a na l yti xl a bs .co.i n Email: i nfo@a na l yti xl a bs .co.i n phone: +91-88021-73069



"srv\_diff\_host\_rate","dst\_host\_count","dst\_host\_srv\_count","dst\_host\_same\_srv\_rate", "dst\_host\_diff\_srv\_rate","dst\_host\_same\_src\_port\_rate", "dst\_host\_srv\_diff\_host\_rate","dst\_host\_serror\_rate","dst\_host\_srv\_serror\_rate", "dst\_host\_rerror\_rate","dst\_host\_srv\_rerror\_rate","attack", "last\_flag"]

## BASIC FEATURES OF EACH NETWORK CONNECTION VECTOR

1. **Duration:** Length of time duration of the connection
2. **Protocol\_type:** Protocol used in the connection
3. **Service:** Destination network service used
4. **Flag:** Status of the connection – Normal or Error
5. **Src\_bytes:** Numberof data bytes transferred from source to destination in single connection
6. **Dst\_bytes:** Numberof data bytes transferred from destination to source in single connection
7. **Land:** if source and destination IP addresses and port numbers are equal then, this variable takes value 1 else 0
8. **Wrong\_fragment:** Total numberof wrong fragments in this connection
9. **Urgent:** Numberof urgent packets in this connection. Urgent packets are packets with the urgent bit activated

## CONTENT RELATED FEATURES OF EACH NETWORK CONNECTION VECTOR

1. **Hot:** Numberof „hot‟ indicators in the content such as: entering asystem directory, creating programs

and executing programs

1. **Num\_failed \_logins:** Count of failed login attempts
2. **Logged\_in Login Status**: 1 if successfully logged in; 0 otherwise **13 Num\_compromised**: Numberof ``compromised'' conditions **14 Root\_shell:** 1 if root shell is obtained; 0 otherwise
3. **Su\_attempted:** 1 if ``su root'' command attempted orused; 0 otherwise
4. **Num\_root:** Numberof ``root'' accesses or numberof operations performed as a root in the connection
5. **Num\_file\_creations:** Numberof file creation operations in the connection
6. **Num\_shells:** Numberof shell prompts
7. **Num\_access\_files**: Numberof operations on access control files
8. **Num\_outbound\_cmds:** Numberof outbound commands in an ftp session **21 Is\_hot\_login:** 1 if the login belongs to the ``hot'' list i.e., root or admin; else 0 **22 Is\_guest\_login:** 1 if the login is a ``guest'' login; 0 otherwise

## TIME RELATED TRAFFIC FEATURES OF EACH NETWORKCONNECTION VECTOR

1. **Count:** Numberof connections to the same destination host as the current connection in the past two seconds
2. **Srv\_count:** Numberof connections to the same service (port number) as the current connection in the past two seconds
3. **Serror\_rate:** The percentage of connections that have activated the flag (4) s0, s1, s2 or s3, among the connections aggregated in count (23)
4. **Srv\_serror\_rate:** The percentage of connections that have activated the flag (4) s0, s1, s2 or s3, among

the connections aggregated in srv\_count (24)

1. **Rerror\_rate:** The percentage of connections that have activated the flag (4) REJ, among the connections aggregated in count (23)
2. **Srv\_rerror\_rate:** The percentage of connections that have activated the flag (4) REJ, among the connections aggregated in srv\_count (24)
3. **Same\_srv\_rate:** The percentage of connections that were to the same service, among the connections aggregated in count (23)
4. **Diff\_srv\_rate:** The percentage of connections that were to different services, among the connections aggregated in count (23)

AnalytixLabs, Website: www.a na l yti xl a bs .co.i n Email: i nfo@a na l yti xl a bs .co.i n phone: +91-88021-73069



1. **Srv\_diff\_host\_ rate:** The percentage of connections that were to different destination machines among the connections aggregated in srv\_count (24)

## HOST BASED TRAFFIC FEATURES IN A NETWORK CONNECTION VECTOR

1. **Dst\_host\_count:** Numberof connections having the same destination host IP address
2. **Dst\_host\_srv\_ count:** Numberof connections having the same port number
3. **Dst\_host\_same \_srv\_rate:** The percentage of connections that were to the same service, among the connections aggregated in dst\_host\_count (32)
4. **Dst\_host\_diff\_ srv\_rate:** The percentage of connections that were to different services, among the connections aggregated in dst\_host\_count (32)
5. **Dst\_host\_same \_src\_port\_rate**: The percentage of connections that were to the same source port, among the connections aggregated in dst\_host\_srv\_count (33)
6. **Dst\_host\_srv\_ diff\_host\_rate:** The percentage of connections that were to different destination machines, among the connections aggregated in dst\_host\_srv\_count (33)
7. **Dst\_host\_serro r\_rate: The** percentage of connections that have activated the flag (4) s0, s1, s2 or s3, among the connections aggregated in dst\_host\_count (32)
8. **Dst\_host\_srv\_s error\_rate:** The percent of connections that have activated the flag (4) s0, s1, s2 or s3, among the connections aggregated in dst\_host\_srv\_count (33)
9. **Dst\_host\_rerro r\_rate**: The percentage of connections that have activated the flag (4) REJ, among the connections aggregated in dst\_host\_count (32)
10. **Dst\_host\_srv\_r error\_rate:** The percentage of connections that have activated the flag (4) REJ, among the connections aggregated in dst\_host\_srv\_count (33)

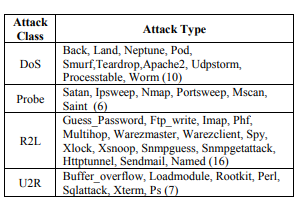
## Type Features:

**Nominal:** Protocol\_type(2), Service(3), Flag(4)

**Binary:** Land(7), logged\_in(12), root\_shell(14), su\_attempted(15), is\_host\_login(21),, is\_guest\_login(22)

**Numeric:** Duration(1), src\_bytes(5), dst\_bytes(6), wrong\_fragment(8), urgent(9), hot(10), num\_failed\_logins(11), num\_compromised(13), num\_root(16), num\_file\_creations(17),

num\_shells(18), num\_access\_files(19), num\_outbound\_cmds(20), count(23), srv\_count(24), error\_rate(25), srv\_serror\_rate(26), rerror\_rate(27),srv\_rerror\_rate(28), same\_srv\_rate(29),diff\_srv\_rate(30), srv\_diff\_host\_rate(31), dst\_host\_count(32), dst\_host\_srv\_count(33), dst\_host\_same\_srv\_rate(34), dst\_host\_diff\_srv\_rate(35), dst\_host\_same\_src\_port\_rate(36), dst\_host\_srv\_diff\_host\_rate(37), dst\_host\_serror\_rate(38), dst\_host\_srv\_serror\_rate(39), dst\_host\_rerror\_rate(40), dst\_host\_srv\_rerror\_rate(41)



## ATTACK CLASS:

AnalytixLabs, Website: www.a na l yti xl a bs .co.i n Email: i nfo@a na l yti xl a bs .co.i n phone: +91-88021-73069



1. **DOS:** Denial of service is an attack category, which depletes the victim‟s resources thereby making it unable to handle legitimate requests – e.g. syn flooding. Relevant features: “source bytes” and “percentage of packets with errors”
2. **Probing:** Surveillanceand otherprobing attack‟s objective is to gain information about the remote victim

e.g. port scanning. Relevant features: “duration of connection” and “source bytes”

1. **U2R:** unauthorized access to local superuser(root) privileges is an attack type, by which an attackeruses a normal account to login into a victim system and tries to gain root/administratorprivileges by exploiting some vulnerability in the victim e.g. bufferoverflow attacks. Relevant features: “numberof file creations” and “numberof shell prompts invoked,”
2. **R2L:** unauthorized access from aremote machine, the attackerintrudes into aremote machine and gains local access of the victim machine. E.g. password guessing Relevant features: Network level features – “duration of connection” and “service requested” and host level features - “numberof failed login attempts”

AnalytixLabs, Website: www.a na l yti xl a bs .co.i n Email: i nfo@a na l yti xl a bs .co.i n phone: +91-88021-73069