International Journal of Emerging Engineering Research and Technology
Volume 2, Issue 2, May 2014, PP 78-85

On Dynamic Rule Generation for Real Time IDS Using GABIDS

Ms.Shital Nagarare Dr. A.M.Dixit
PG Student, Department of Computer Engineering
PVPIT, Bavdhan, Pune, India

Abstract: *GABIDS (**G**enetic **A**lgorithm **B**ased **I**ntrusion **D**etection **S**ystem (IDS)) is a dynamic IDS to support the real-time rule generation to identify abnormal network behavior. The chromosomes for the GA are essentially produced from information connectivity in the network. The currently existing IDS typically detect the attack type on basis of traditional static methods like fingerprinting. These IDS may not be robust enough to handle a new attack-type. In this paper, we propose GABIDS framework which dynamically generates the new set of rules to handle the unknown abnormal network behavior at any instant and update the Rule-database accordingly. This feature dynamically updates the rule set by capturing each new and unique network behavior state. GABIDS uses KDD Cup 99 dataset for training and testing purpose.*

Index Terms: *Network intrusion detection system, genetic algorithm, network security, KDD dataset, CRN, MANET.*

1. INTRODUCTION

The Information Technology has become a fundamental necessity of our lives today. We cannot imagine a day without rendering a service from computer based system(s). As a coin has two sides, these services too come with some challenges. As a result of these services it also increases the issue of commuter service. With every passing day, the Prevention and recognition of intrusion to computer technology is becoming more and more challenging. The number of rising threats has given rise to a new arena of Cyber security. Network security is the new ground of learning to handle security issues in general.

A variety of security mechanisms are designed to enforce and support legitimate activities while handling the system resources and data. However, it appears that such mechanisms are not able to completely prevent intrusions into computer systems. It is essential to identify the intrusion correctly, so that impact of intrusion and damage can be restored. Likewise, effective detection of the intrusion will aid security experts to advise measures that can be used to prevent them from happening in the future. Intrusion detection system having the capability to control different type of attack for this purpose there are various methodology and approaches has been developed to computer systems.

Dr. Dorothy Denning proposed a model in 1987 which is known as intrusion detection system, which became a landmark in the research in computer security area. She proposed the fundamental concept which forms core of most intrusion detection methodologies in use today [19].

The genetic algorithm is inspired by biological concept like inheritance, mutation, selection, and crossovers. The genetic algorithm is based on “Darwinian principle of Evolution”, in working, which advocates among a population chosen, the survival of the fittest [20][21].

Hence, a result acquired by introducing genetic algorithms to any problem consists of only those best solutions which are said to fulfil a predefined fitness value. The simulated attack falls in one of the following four categories: [22] [9]:

1. **Denial of Service Attack (DOS**): In this category the attacker makes some computing or memory resources too busy or too full to handle legitimate request, or deny legitimate users access to machine. DOS contains the attacks: 'neptune', 'back', 'smurf', 'pod', 'land', and 'teardrop'.
2. **Users to Root Attack (U2R):** In this category the attacker starts out with access to a normal user account on the system and is able to exploit some vulnerability to obtain root access to the system. U2R contains the attacks: 'perl', 'loadmodule', 'rootkit' and 'buffer_overflow'
3. **Remote to Local Attack (R2L):** In this category the attacker sends packets to machine over a network but who does not have an account on that machine and exploits some vulnerability to gain local access as a user of that machine. R2L contain the attacks: 'spy' and 'phf' , '

ftp_write' 'guess_passwd' 'warezclient', ' multihop', , 'imap', 'warezmaster',

4. **Probing Attack (PROBE)**: In this category the attacker attempt to gather information about network of computers for the apparent purpose of circumventing its security. PROBE contains the attacks: 'portsweep', 'satan', 'nmap', and 'ipsweep'

Existing System and Their Drawbacks

A. Existing System

In early work on Security, Salter and Schroeder established security design principal and mechanisms as well as Orange Book design for DOD (Department of Defense) specification. The examples for early system are IDES used for statically anomaly-detection, Haystack added signature detection, and ISOA uses both real-time monitoring and post-session analysis to detect doubtful behavior, established summaries at both levels

B. Drawbacks of existing system

- Manual energy required to conquer the behavior and flow of packets.
- They work on traditional method like "fingerprinting" means the signature base, if the signature is match then it will detect the malicious activity otherwise not.
- They cannot create the rule at runtime.
- They cannot take the right decision with respect to the current situation.
- They only work on initial static rule set applied so this results in lack of ability to decision-making during new unknown kind of packets.

2. LITERATURE REVIEW

Combating Attacks against Cognitive Radio Networks (CRN) providing spectrum scarcity solution as they are susceptible security threats. This describes CRN based Wireless Regional Area Network (WRAN) and its security threats. This uses non-parametric cumulative sums (CUSUM) to detect suspectful behavior of system due to attack. This adopts Anomaly Detection and provides to CRN system parameters through a learning phase. The proposed IDS are evaluated through computer based simulations, and the simulation results clearly indicate the effectiveness of this proposal [4].

IDS is a key element to system security that monitors susceptible events in system or network. Among various IDS methodologies no one methodology is ideal. The attack prediction system may be generating false alarm which uses anomaly detection system. By using fizzy logic false alarm rate can be reduced. With fizzy logic some strategy is needed to best detection result of abnormal behavior in system or network [5].

MANET is formed using multiple wireless mobile devices without any centralized administration. Every device can communicate with other devices which are in its radio communication range with wireless trans-receiver. In MANET no communication taken place with device which out of its radio range. Each wireless node act as host and router at same time. In MANET IDS prevents vulnerability attack. Intrusion Prevention is the primary defense [6].

Signature based IDS collects information from log & events that information is used to provide security to network. Signature based IDS methodologies are important to protect target systems and networks. Signature based IDS is most extensive threat detection technique. It cannot manage traffic flood so this may drop potential attacks. The Signature based IDS using mobile agents, detects threats with extremely high success rate dynamically and automatically created small and efficient multiple databases and provides mechanism to update signature databases in particular period of time [7].

Here they used NSL-kdd dataset with Weka mining tool as input. In NSL –kdd there are25192 classified instances with 41 attribute and out of these 12 are selected for evaluation purpose using weka tool. Four data mining Neural Classifier algorithms are used to test the dataset like, Voted Perception, Logistic, Multilayer Perception, and RBF Network. Kappa is a chance-corrected measure of agreement between the classification and the true classes [8].

3. INTRUSION DETECTION SYSTEM (IDS)

Today's internet system of working creates more security issues for a Single system or for a Network based systems, Like Hacking, Bot attacks, Spyware attacks, Virus attacks or Malware attacks. IDS provide security against suspicious inbound and outbound traffic.

IDS, is a security system Device or Software that used against unauthorized access or intrusion. IDS is not an active process, because IDS doesn't monitor event till attack is taken

place. IDS works when detect breach in security thread in a Host or a Network by monitoring events that are taken place

NID has traditionally been unable of working in the following environment:

- Switch network
- Incepted network
- High speed network (anything over 100mbps)

Recently however cisco release a module for their catalyst 6000 switch that incorporate network intrusion detection directly in the switch, disabling all these flaws. Also, ISS/network ice indicates that they are now capable of "packet sniffing" at gigabit speed.

IDS' working is classified into two types as follows:

A. Signature Based Intrusion Detection System

In this method every data packet is checked by attack definition stored in attack database. If it found any packet matched with definition pattern signature based IDS generate a report message for administrator by E-mail, SMS or any other communication device. Also known as Knowledge based IDS.

B. Anomaly based Intrusion Detection System

In Anomaly based IDS monitor network traffic and analyze through administrator's security baseline. In such security administrator define rules for bandwidth allocation for each resource, Protocols, Ports allowed for users and devices communicating in network. Also, says as Behavior based IDS.

3.1. Types of IDS

Following are the example of IDS that are being used now days as follows:

- HIDS
- NIDS
- Host based Intrusion Detection System (HIDS)

- HIDS monitor inbound and outbound traffic of single host which is connected in network.
- HIDS uses OS monitor tools like event viewer and log events generated by each user.
- Analyze user shell commands entered by user.
- Monitor sends mail events.
- System registry and system calls.

In HIDS, it monitors the entire system (host) and check the each packets/data available in the host

- *Network based Intrusion Detection System (NIDS)*

- NIDS is single hardware application that placed at network gateway along with firewall that monitors inbound and outbound network traffic.
- NIDS placed hardware sensors at various network locations like Routers and Proxy servers.
- Monitor violated protocols like TCP/UDPbased and ports access by any user or terminal.
- NIDS detects insertion and evasion type of attacks.
- Firewall is active filtering while NIDS is passive filtering.
- NIDS requires High bandwidth and real-time alerts

4. GENETIC ALGORITHM

The *genetic algorithm* is a probabilistic search algorithm that iteratively transforms a set of mathematical objects, each with are related fitness value, into a new population of offspring objects using the Darwinian principle of natural selection and using operations that are patterned after naturally occurring genetic operations, such as crossover, inheritance, selection and mutation. These algorithms convert a potential solution to a specific problem on a simple chromosome like data structure and apply recombination operators to these structures so as to preserve critical information. Genetic algorithms are often observed as function optimizers although the range of problems to which genetic algorithms have been applied is quite comprehensive. Genetic Algorithms are good at taking large, potentially enormous search spaces and piloting them, looking for best combinations of things, solutions you might not otherwise find in a lifetime.

GAs differs from other heuristic methods in several ways. One important difference is that it works on a population of possible solutions, which other heuristic methods use a single

solution in their iterations. Another difference is that GAs is probabilistic and not deterministic.

4.1. Outline of the Genetic Algorithm

1. Generate a random population of n chromosomes which are suitable solutions.
2. Establish a method to evaluate the fitness f(x) of each chromosome x in the population.
3. Create a new population by repeating the following steps until the new population is complete
   - Selection - select from the population according to some fitness scheme.
   - Crossover- New offspring formed by a crossover with the parents
   - Mutation - With a mutation probability mutate new offspring at each locus (position in chromosome).
4. Use the newly generated population for a further run of algorithm.

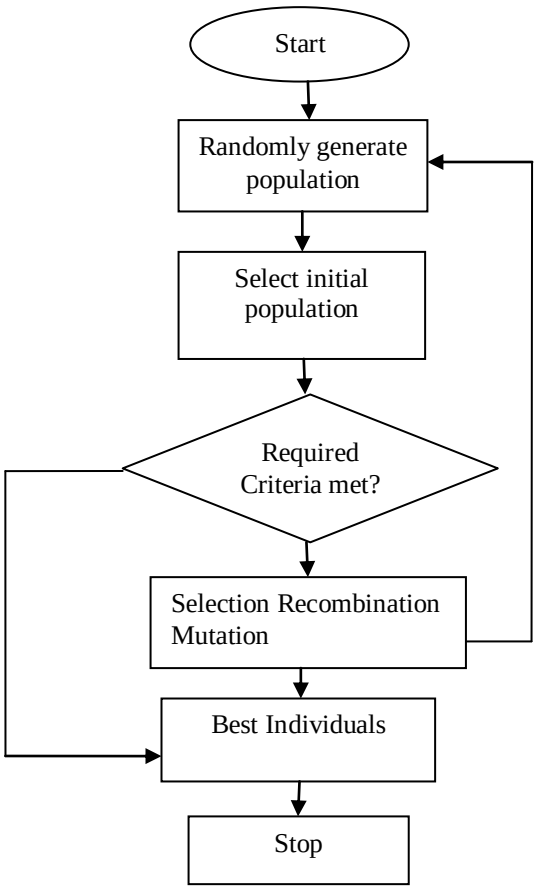

Fig 2. *Genetic Algorithm*

5. KDD CUP 99

To view different areas of IDS, in Lincoln Laboratory at MIT under the Defense Advanced Research Project Agency(DARPA) and Air Force Research Laboratory(AFRL/SNHS) helping to construct the first standard dataset for NIDS. It has two phases one is training dataset and testing dataset. There are 41 attributes are available [10].

6. PROPOSED SYSTEM: GABIDS

As a result of growth in the computer technology growth it is imperative to conserve a high level security to keep a trusted communication over the internet or various network places. But secured data communication over internet and any other network is always under threat of intrusions and misuses. So Intrusion Detection Systems have become a needful component in terms of computer and network security. Here we present Network Intrusion Detection System (NIDS), by applying genetic algorithm (GA) to efficiently detect various types of network intrusions. System creates a new set of rules during run time. So the intruder cannot be able to attack the system with virus. The Intrusion Detection System in Networking Using Genetic Algorithm (IDS) is to identify the intruder and block the data from the intruder to avoid the system attack by the virus. The major components of the system are creating new set of rules during run time. We used weka mining tool that is capable to to extract the needed data. The data is partitioned into the two classes of "Normal" and "Attack" patterns where Attack is the collection of four classes (Probe, DOS, U2R, and R2L) of attacks. And our implementation is to determine the subtype of a particular attack. This new system is a replacement of the existing system. To implement and measure the performance of our system we used the KDD99 benchmark dataset and obtain reasonable detection rate[8].

A. GABIDS Architecture

By using Audit trail data the new chromosomes will be generated. Audit trail data being feed with record sets like Destination & Source hardware and logical addresses, logical ports, flags and etc. These chromosome sets are being manipulated by Genetic algorithm and send to crossover and mutation that will generate the population. The newly generated population is compared with training dataset to measure fitness function.

50% Selection criteria is used on best selected individuals and remaining will be either again processed with GA or discarded. These selected individuals age chromosomes which are used as real-time generated rules set. These new rules set are upgraded to testing dataset. The aim is to classified attacks and its patterns also detect its subtypes. These newly generated rules compare signature of every packet with attack definition stored in testing dataset. And signature matched it will detect as an attack

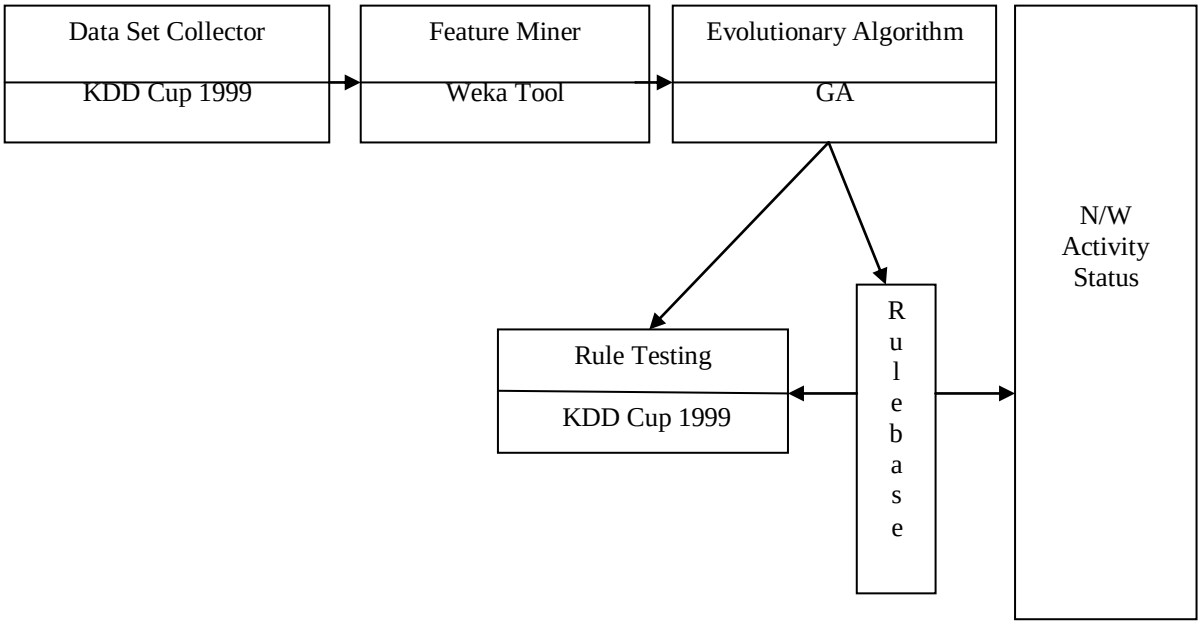

Fig. 3: *GABIDS System Architecture*

B. GABIDS Algorithm

The proposed GABIDS algorithm consists of the following steps:

Step 1: Datasets Collector

Collect the rules form KDD Cup 99(Training Dataset)

Step 2: Feature Mining

Extract features using Weka tool

Source, flag, protocol, Dst_Byte, Src_Byte, Duration

Step 3: Evolutionary Algorithm

Used GA with three parameter

1) Selection

2) Crossover

3) Mutation

Generate New Population

Step 4: Testing Rule (KDD Cup 99)

0, ftp_data, SF, 491, 0, normal

0, udp, other, SF, 146, 0, normal

0, tcp, private, S0, 0, 0, anomaly

Step 5: Extracted Rule

New rule will be added here

Step 6: Network Activity Status

Normal

Anomaly

C. GABIDS Implementation

Implementation part is most crucial part to succeed a successful system and to give user confidence over the new system that is workable and achievable.

1. Datasets Collector

Training phase cover individually rules for each with anomaly connection as well as normal connection of the network. Training dataset consists of approximately 4900000 single connection vectors each of which contains 41 features and is labeled as either normal or an attack, which exactly one specific attack type[9] [11].

2. Weka Tools

In previous surviving system has been established with various outfits. In this case, we extract the desirable attribute from KDD dataset by using only Weka tool [12].

Feature name	Description	Format	Type	# of genes
duration	length (number of seconds) of the connection	H:M:S	continuous	1
Protocol type	type of the protocol, e.g. tcp, udp, etc.	String	discrete	1
Service	network service on the destination, e.g., http, telnet, etc.		discrete	1
src_bytes	number of data bytes from source to destination	Numeric	continuous	1
dst_bytes	number of data bytes from destination to source	Numeric	continuous	1
flag	normal or error status of the connection		discrete	1
Class	Attack name and type	String	continuous	1

Training data contains connection such that it known apriori, connections that are attacks.Training phase uses a subset of DARPA dataset, which contains all 7 features so that a connection considered intrusive, has an attack name. The data analysis prior to its use notes that normal connection in the training data contains no attack name. Each chromosome is a rule within which the 7 features are encoded via fixed length, and each features encoded as one or more genes of different types as in table.

3. Evolutionary Algorithm

Step 1: Generate chromosome

Set max_gen_val,min_gen_val

Step 2: Randomly generate the population

Set pop size; var_size

Step 3: crossover rate and mutation rate=0.84

For (int i = 0; i < IndivisualSize; i++)

Mutation if (ran <= mutationRate)

Step 4: calculate fitness

GenerateIndividuals ()

For(i=0;i<pop_size;i++)

{

For (int i = 0; i < IndivisualSize; i++)

End for

If (ran <= mutationRate)

End if

maxFit = (double) (tp + fn) / total

}

End for

Step 5: new generation

Step 6: stop

In the phase of training, from preparation dataset we removed the redundant record to see the effect of the detection rate. Hence Weka tool is use to extract these feature.

Now we require extracted features from training dataset and these is duration, flag src_byte, Dest_byte protocol, service etc.

The composed attributed are converted into the chromosomes contained by the range and in the same behavior. The genetic algorithm starts with randomly generated population of chromosome.

If the available chromosome does not match with the training dataset, it will be considered as a new rule. Further it will be added to the database if it does not match with the fitness function, so it will go again for mutation, selection and crossover function of genetic algorithm.

Crossover: crossover function having two parameters known as chromosome, so we consider those parameters as parent chromosome. And the crossover rate is 0.84. After crossover we are getting the new chromosome as well as an old chromosome.

Mutation: Now we have old and new population and our mutation rate is 0.86. Here mutation function will change or flip any random bit of chromosome. This newly

generated population will compare with each rule which is available in training dataset of kdd99 on the basis of fitness function

Selection: The available value for each chromosome will be selected on the basis of 50% suitable criteria and the remaining ones are removed.

4. Rule Testing

In testing dataset for GABIDS, its rules are checked with the rules which are generated by Genetic Algorithm. If existing rules match with GA rules then it will generate the signal of intrusion, conveying that this is anomaly connection. If existing rules do not match with GA rules, it is considered to be a new rule and then added to the rule base.

7. GABIDS ADVANTAGES

Following are the advantages of proposed GABIDS framework:

- It eliminates the need for an attack to be previously known t for successful attack detection, because malicious behavior is different from normal behavior in nature.
- Using a generalized behavioral model is theoretically more accurate, efficient and easier to maintain than a fingerprinting system.
- It uses a constant amount of computer resources per user, reducing the possibility of depleting available resources.
- Once installed, there is no need for constant human assistance and intervention to monitor the system.
- It generates its own rules depending on the real-time behavior of the packet transmission.
- It dynamically increases the rules in the dataset according to the packets flowing in the network and the decisions taken by the system. Due to the increase of rules in the rule set, the reliability and robustness of the system also increases.

8. CONCLUSION

GABIDS is a modern system security technique adding more advanced features to current firewalls and router security. In current scenario the system and data security is one of the most challenging tasks in networking environment. Genetic Algorithm based IDS uses GA which is based on mathematical model performing calculations of population of chromosomes such as fitness function of genes. GA depends on Selection, Crossover and Mutation methods to generate new rules for population with testing data set.

GABIDS is capable of a real time rule generation for known as well as unknown network activity in the system. To implement and measure the performance of our system we used kdd dataset with weka (mining) tool to obtain reasonable detection rate. The existing GABIDS framework can be made more robust as an Intrusion detection system with the help of sophisticated statistical analysis methodology.

REFERENCES

[1] Mohammad SazzadulHoque, Md. Abdul Mukit and Md. Abu NaserBikas ”An Implementation of Intrusion Detection System using Genetic Algorithm”, International Journal of Network Security & Its Applications (IJNSA), Vol.4, No.2, March 2012.

[2] EmadElbeltagi, TarekHegazy, Donald Grierson,” Comparison among five evolutionary-based optimization algorithms”2005.

[3] A .Sung , S Mukkamala,” Identifying important features for intrusion detection system using support vector machine and neural networks”in Symposium on Application and the Internet,pp. 209-216.2003.

[4] Zubair Md. Fadlullah, Hiroki Nishiyama, and Nei Kato,” An Intrusion Detection System (IDS) for Combating Attacks against Cognitive Radio Networks.”Tohoku University Mostafa M. Fouda, Tohoku University and Benha University(2013IEEE)

[5] Current studies on intrusion detection system, genetic algorithm and fuzzy logic. Mostaque Md. Morshedur Hassan LCB College, Maligaon, Guwahati, Assam, India.(march 2013 International distributed and parallel sytem)

[6] Treesa Nice P. A.Department of Computer Science and Engineering “A Survey on Intrusion Detection Systems in Mobile Ad Hoc Networks “, Rajagiri School of Engineering and Technology, Ernakulam, India(may 2013 IJRCCE)

[7] Mrmueenuddin, Mr. azizah ADDUAL rEHMAN. “Signature-based Multi-Layer Distributed Intrusion Detection System using Mobile Agents” Jan 2013International journals of network security.

[8] Gong R.H, Zulkemine.M, Anolmaesumi.P, "A Software Implementation of a Genetic Algorithm Based approach to Network Intrusion Detection", Proceedings of the SNPD/SAWN'05, PP.19-27, Aug 2005.

[9] Adel NadjaranToosi , Mohsen Kahani , "A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers (2007).

[10] MIT Lincoln Labs, 1998 DARPA Intrusion Detection Evaluation. Available on: http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/index.html, February 2008.

[11] wekatools;http://www.ns5.org

[12] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[13] Sectools.Org: 2006 Results; http://sectools.org/tools2006.html.

[14] SecTools.Org: Top 125 Network Security Tools; http://sectools.org/tag/ids/

[15] Snort (software); http://en.wikipedia.org/wiki/Snort_%28software%29.

[16] InfoWorld, The greatest open source software of all time, 2009; http://www.infoworld.com/d/open source/greatest-open-source-software-all-time- 776? Source= fssr.

[17] http://en.wikipedia.org/wiki/Suricata_ (software).

[18] The Bro Network Security Monitor; http://bro-ids.org/

[19] Denning, Dorothy. (February, 1987). An Intrusion-Detection Model. IEEE Transactions on Software Engineering, Vol. SE-13, No. 2.] In the middle of 1987.

[20] F. EidHebba, A. Darwish, A. E. Hassanien, and K. Tai-Hoon, "Intelligent Hybrid Anomaly Network Intrusion Detection System," in CCIS 265, 2011, vol. Part I, pp. 209–218.],

[21] K. G. Srinivasa, "Application of Genetic Algorithms for Detecting Anomaly in Network Intrusion Detection Systems," Lecture Notes of The Institute for Computer Sciences, Social Informatics & Telecommunication Engineering, vol. 84, pp.582–591, 2012

[22] M. Tavallaee, E. Bagheri, W. Lu and A. Ghorbani, "A Detailed Analysis of the KDD'99 CUP Data Set", The 2nd IEEE Symposium on Computational Intelligence Conference for Security and Defense Applications (CISDA), (2009).