A SURVEY ON ANOMALY-BASED NETWORK INTRUSION DETECTION SYSTEMS

Ankita Choubey*

NaviSinghThakur*

Abstract:

The importance of network security has grown tremendously and a number of devices have been introduced to improve the security of a network. Network Intrusion Detection Systems (NIDS) are among the most widely deployed such system. Popular NIDS use a collection of signatures of known security threats and viruses, which are used to scan each packet's payload.Most IDSs lack the capability to detect novel or previously unknown attacks.A special type of IDS,called Anomaly Detection Systems ,develop models based on normal system or network behaviour,with the goal of detecting both known and unknown attacks. Anomaly detection systems face many problems including high rate of false alarm, ability to work in online mode,and scalability. This paper presents a selective survey of incremental approaches for detecting anomaly in normal system and network traffic.

Keywords: **Computer networks,Network security,Anomaly detection,Intrusion detection**

* **Shree ram institute of science & technology Jabalpur (m.p)**

INTRODUCTION

The field of intrusion detection has received increased attention in recent years .Onereason for this is the explosive growth of the Internet and the large number ofnetworkedsystemsthatexist in all types of organizations. The in crease in the number of networked machines has lead to an increase in unauthorized activity, not only from externalattackers,butalsofrominternalattackers, such as disgruntled employee and people abusingtheirprivilegesforpersonalgain.

Securityisabigissuefor allnetworksintoday's enterprise environment. Hackers and intruders have made many successful attempts to bring down high-profile company networks and web services. Many methods have been developed to secure the network infrastructure and communication over the Internet, among them theuseoffirewalls, encryption,andvirtualprivate networks. Intrusion detection is a relatively new addition to such techniques. Intrusion detection methodsstartedappearinginthelastfewyears. Using intrusion detection methods, you can collect and use

information from known types of attacks and findout if someone is trying to attack your network or particular hosts.The information collected this way can be used to harden your network security, as well as for legal purposes. Both commercial and open source products are now available for this purpose.Many vulnerability assessment tools are also available in the market that can be used to assess different types of security holes present in your network.

CLASSIFICATION OF INTRUSION DETECTION SYSTEM

Alltheclassificationofintrusiondetectionsytem is described below as shown in Figure1.

StatisticalModels

Operational Model/ ThresholdMetric

The count of events that occur over a period of time determines the alarm to be raised if fewer then"m"ormorethan"n"eventsoccur.Thiscan be visualized in Win2k lock, where a user after "n" unsuccessful login attempts here lowerlimit

is“0”andupperlimitis“n”.Executablefilessize downloaded is restricted in some organizations about 4 MB. The difficulty in this sub-model is determiningmandn(Faizal*etal*.,2009).

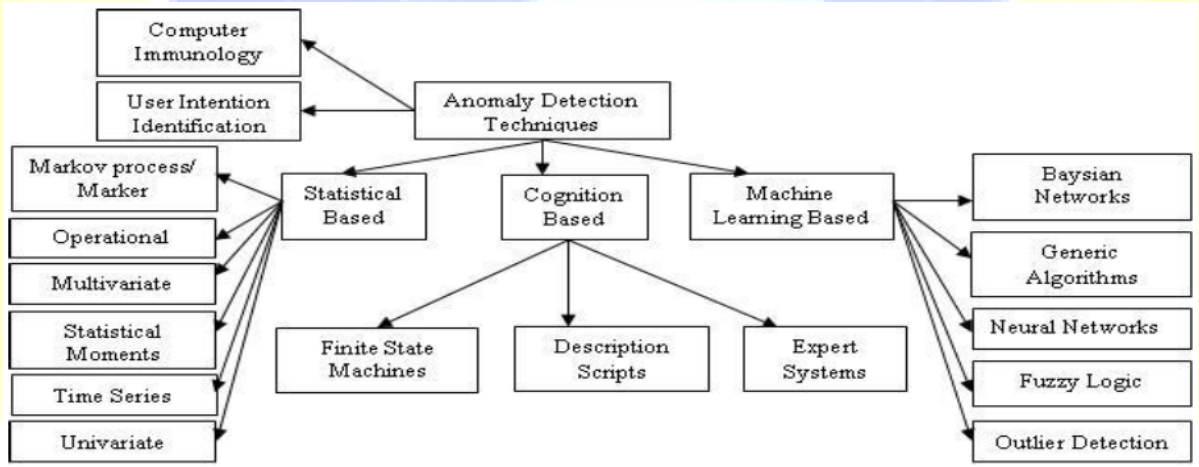
Markov Process or MarkerModel

The Intrusion detection in this model is done by investigating the system at fixed intervals and keeping track of its state a probability for each state at a given time interval *Is*. The change of the state of the system occurs when an event happensandthebehaviorisdetectedasanomaly iftheprobabilityofoccurrenceofthatstateislow. The transitions between certain commands determine the anomaly detection where command sequences wereimportant.

Statistical Moments or Mean and Standard

Deviation Model: In statistical mean, standard deviation,oranyothercorrelationsareknownas a moment. If the event that falls outside the set intervalaboveorbelowthemomentissaidtobe anomalous. The system is subjected to change by considering the aging data and making

Figure1:Classification of Intrusion Detection System



changes to the statistical rule data base. There are two major advantages over an operational model. First, prior knowledge is not required determiningthenormalactivityinordertosetlimits; Second, determining the confidence intervals dependsonobserveduserdata,asitvariesfrom usertouser.Thresholdmodel(Faizal*etal*.,2009) lacksthisflexibility.Themajorvariationonthemean and standard deviation model is to give higher weights for the recentactivities.

Multivariate Model: The major difference betweenthemeanandstandarddeviationmodel is based on correlations among two or more metrics. If experimental data reveals better judicious power can be achieved from combinations of related measures rather than treating themindividually.

Time SeriesModel

Intervaltimerstogetherwithaneventcounteror resourcemeasurearemajorcomponentsinthis model. Order and inter-arrival times of the observationsaswellastheirvaluesarestored.If theprobabilityofoccurrenceofa newobservation is too low then it is considered as anomaly. The disadvantage of this model is that it is more computationallyexpensive.

CognitionModels

Finite StateMachine

AFiniteStateMachine(FSM)orfiniteautomation is a model of behavior captured in states, transitions and actions. A state contains information about the past, i.e., any changes in the input are noted and based on it transition happens.Anactionisadescriptionofanactivity thatistobeperformedatagivenmoment.There areseveralactiontypes:entryaction,exitaction, and transitionaction.

DescriptionScripts

Numerous proposals for scripting languages, which can describe signatures of attacks on computers and networks, are given by the Intrusion Detection community. All of these scriptinglanguagesarecapableofidentifyingthe sequences of specific events that are indicative ofattacks.

AdeptSystems

Human expertise in problem solving is used in adept systems. It solves uncertainties where generally one or more human experts are consulted.Thesesystemsareefficientincertain problemdomain,andalsoconsideredasaclass of Artificial Intelligence (AI) problems. Adept Systems are trained based on extensiveknowledge of patterns associated with known attacksprovidedbyhumanexperts.

Cognition Based Detection Techniques Cognition-Based (also called knowledge-based or expert systems) Detection Techniques work
on the audit data classification technique, influencedbysetofpredefinedrules,classesand attributes identified from training data, set of classification rules, parameters and procedures inferred.

Boosted Decision Tree

BoostedTree(BT),thatusesADABoostalgorithm to generate many Decision Trees classifiers trained by different sample sets drawn from the originaltrainingset,isimplementedinmanyIDS successfully.Allhypotheses,producedfromeach of these classifiers, are combined to calculate total learning error, thereby arriving at a final compositehypothesis.

Support Vector Machine

Support Vector Machines (SVM), reliable on a range of classification tasks, are less prone to over-fittingproblem,andareeffectivewithunseen data. The basic learning process of the SVM includes twophases:(1)Mappingthetrainingdata from the original input space into a higher dimensional feature space, using kernels to transform a linearly non separable problem into a linearly separable one, (2) Finalizing a hyper plane within the feature space, with a maximum margin using Sequential Minimal Optimization (SMO) or Osuna'smethod.

Artificial NeuralNetwork

ArtificialNeuralNetwork(ANN)architecturesAlex Lam (2005) (popular one being , Multilayer Perceptron (MLP), a layered feed-forward topology in which each unit performs a biased weighted sum of their inputs and pass this activation level through a transfer function to

produce their output), are able to identify not readily observable patterns, however MLP is ineffective with new data. For general signal processing and pattern recognition problems, another branch of ANN that makes use ofradial basis function, called The Modified Probabilistic NeuralNetwork(DBarbara*etal*.,2006)(related to General Regression Neural Network(GRNN) classifier and generalization of Probabilistic Neural Network (PNN)), was introduced by Zaknich. It assigns the clusters of input vectors ratherthaneachindividualtrainingcasetoradial units.

Machine Learning Based Detection Techniques

Machine learning techniques Reuters News Service(2005)todetectoutliersindatasetsfrom a variety of fields were developed byGardener

(use a One-Class Support Vector Machine (OCSVM)todetectanomaliesinEEGdatafrom epilepsy patients) and Barbara (proposed an algorithm to detect outliers in noisy datasets wherenoinformationisavailableregardingground truth, based on a Transductive Confidence Machine(TCM))JMaandSPerkins(2003).Unlike induction that uses all data points to induce a model, transduction, an alternative, uses small subsetofthemtoestimateunknownattributesof testpoints.Toperformonlineanomalydetection ontimeseriesdatain,MaandPerkinspresented analgorithmusingsupportvectorregression.Ihler *et al.* present an adaptive anomaly detection algorithm that is based on a Markov-modulated Poisson process model, and use Markov Chain MonteCarlomethodsinaBayesianapproachto learnthemodelparametersAIhler(2006).

Common Attacks and Vulnerabilitiesand Role ofNIDS

Current NIDSs requires substantial amount of human intervention and administrators for an effective operation. Therefore it becomes important for the network administrators to understandthearchitectureofNIDS,andthewell known attacks and the mechanisms used to detect them and contain the damages. In this section, we discuss some well known attacks, exploits, and vulnerabilities in the end host operating systems, andprotocols.

ATTACKTYPES

Confidentiality: In such kinds of attacks, the attacker gains access to confidential and otherwise inaccessibledata.

Integrity: In such kinds of attacks, the attacker can modify the system state and alter the data withoutproperauthorizationfromtheowner.

Availability: Insuchkindsofattacks,thesystem is either shut down by the attacker or made unavailable to general users. Denial of Service attacksfallintothiscategory.

Control: In such attacks the attacker gains full control of the system and can alter the access privileges of the system thereby potentially triggeringalloftheabovethreeattacks.

Attacks Detected by aNIDS

A number of attacks can be detected by current generationofNIDS.Someofthesearelistedand describedbelow.

ScanningAttack

Insuchattacks,anattackersendsvariouskinds of packets to probe a system or network for vulnerability that can be exploited. When probe packetsaresentthetargetsystemresponds;the responses are analyzed to determine the characteristics of the target system and if there arevulnerabilities.ThusscanningattackAlexLam (2005) essentially identifies a potential victim. Network scanners, port scanners, vulnerability scanners, etc., are used which yields these information. Once the victim is identified, the attacker can penetrate them in a specific way. Scanning is typically considered a legal activity and there are a number of examples and applicationsthatemployscanning.The most well known scanning applications are Web search engines.Ontheotherhandin dependent individual ay scan a network or the entire Internet looking forcertainin formation,suchasamusicor video file.Somewell-known malicious scanning include Verticaland Horizontal port scanning,ICMP(ping) scanning, very slow scan,scanning from multiple portsandscanning of multiple IPad dresses and ports .NIDSsignaturescanbedevisedtoidentify such malicious scanning

activityfromalegitimatescanning activity with fairly high degree of accuracy.

Denial of Service (DoS)Attacks

ADenialofServiceattackattemptstoslowdown or completely shut down atar get so as to disrupt theserviceanddenythelegitimate and authorized users an access. Such attacksar ever ycommon intheInternetwhereacollectionofhostsareoften used to bombard web servers with dummy requests. Such attacks can cause significant economic damage to e-commerce businesses by denying the customers an access to the business. There are a number of different kinds ofDoSattacks(JMaandSPerkins(2003),some of which are mentionedbelow.

Flaw Exploitation DoSAttacks

Insuchattacks,anattackerexploitsaflawinthe serversoftwaretoeitherslowitdownorexhaust it of certain resources. Ping of death attack is one such well known attack. A ping of Death (POD) (Alex Lam 2005) is a type of attack on a computer that involves sending a malformed orotherwisemaliciouspingtoacomputer.Apingis normally 64 bytes in size (or 84 bytes when IP header is considered); many computer systems cannot handle a ping larger than the maximum IPpacketsize,whichis65,535bytes.Sendinga ping of this size can crash the target computer. Some limitations of the protocol implementation also lead to vulnerability which can beexploited toimplementDoSattacksJelenaMirkovic*etal*., (2005) such as DNS amplification attack which usesICMPechomessagestobombardatarget. For these attacks, a signature can be devised easily,suchastodetermineapingofdeathattack aNIDSneedstocheckthepingflagandpacket size.

Flooding DoS Attacks

In a flooding attack, an attacker simply sends morerequeststoatargetthatitcanhandle.Such attacks can either exhaust the processing capability of the target or exhaust the network bandwidth of the target, either way leading to a denialofservicetootherusers.DoSattacksare extremely difficult to combat, as these do not exploit anyvulnerability in the system, and even an otherwise secure system can be targeted. A moredangerousversionofDoSattack(Reuters NewsService(2005),iscalledDistributedDenial ofServiceattack(DDoS),whichusesalargepool of hosts to target a given victim host. A hacker (calledbotmaster)caninitiateaDDoSattackby exploitingvulnerabilityinsomecomputersystem, thereby taking control of it and making this the

DDoS master. Afterwards the intruder uses this master to communicate with the other systems (called bots) that can be compromised. Once a significant number of hosts are compromised, withasinglecommand,theintrudercaninstruct themtolaunchavarietyoffloodattacksagainst a specifiedtarget.

PenetrationAttacks

Inpenetrationattack(AlexLam2005),anattacker gains an unauthorized control of a system, and can modify/alter system state, read files, etc. Generallysuchattacksexploitcertainflawsinthe software, which enables the attacker to install viruses, and malware in the system. The most common types of penetration attacksare:

UsertoRoot: Alocalusergetsthefullaccessto every component of thesystem.

Remote to User: A user across the network gainsauseraccountandtheassociatedcontrols.

Remote to Root: A user across the network gains the complete control of the system.

RemoteDiskRead: Anattackeronthenetwork gains access to the inaccessible files stored locally on thehost.

Remote Disk Write: An attacker onthe networknotonlygainsaccesstotheinaccessible filesstoredlocallyonthehost,butcanalsoalter them.

SSHAttack

SSH attacks are a main area of concern for networkmanagers,duetothedangerassociated with a successful compromise. The fact thatthe numberofpeopleusingandrelyingontheInternet is increasing rapidly makes breaking into and compromising systems an ever more lucrative activity for hackers. One popular class of attack targets is that of Secure Shell (SSH) daemons. BymeansofSSH(AlexLam2005),ahackercan gain access to and potentially full control over remotehosts.Oncecompromised,ahackercan sabotage not only the host itself, but also use it for attacking other systems. The detection of intrusions, especially in the case of SSH, is thereforecrucialforpreventingdamagetohosts andnetworks.

INTRUSION DETECTION SYSTEMS

IntrusionDetectionSystem(IDS)issoftwarethat automates the intrusion detection process and detects possible intrusions. IDS serve three essentialsecurityfunctions:theymonitor,detect, andrespondtounauthorizedactivitybycompany insiders and outsider intrusion. An IDS is composed of severalcomponents:

Sensors (S K Sharma *et al*., 2012) which generate security events; Console to monitor eventsandalertsandcontrolthesensorsCentral Enginethatrecordseventsloggedbythesensors

in a database and uses a system of rules to generate alerts from security eventsreceived.
In many simple IDS implementations (Tarem Ahmed*etal*.,2007)thesethreecomponentsare combined in a single device or appliance. More specifically, IDS tools aim to detect computer attacksand/orcomputermisuse,andtoalertthe properindividualsupondetection.

IDSsusepoliciestodefinecertaineventsthat, ifdetectedwillissueanalert.Inotherwords,ifa particular event is considered to constitute a security incident, an alert will be issued if that eventisdetected.CertainIDSshavethecapability ofsendingoutalerts,sothattheadministratorof the IDS will receive a notification of a possible securityincident Intheformofapage,e-mail,or SNMPtrap(LIYongzhong*etal*.,2008).ManyIDSs notonlyrecognizeaparticularincidentandissue an appropriate alert, they also respond automaticallytotheevent.Sucharesponsemight include logging off a user, disabling a user account, and launching of scripts. IDSs are an integral and necessary element of a complete informationsecurityinfrastructureperformingas "the logical complement to networkfirewalls"

.Simply put, IDS tools allow for complete supervisionofnetworks,regardlessoftheaction being taken, such that information willalways exist to determine the nature of the security incidentanditssource.Ideallytheteam'snetwork is separated from the outside world by a well designedfirewall.Theoutsideworldincludesthe team's host organization. Firewalls protect a networkand attempt topreventintrusions,while IDS tools detect whether or not the network is underattackorhas,infact,beenbreached.IDS toolsthusformanintegralpartofathoroughand complete security system. They don't fully guarantee security, but when used withsecurity

policy, vulnerability assessments, data encryption, user authentication, access control, and firewalls, they can greatly enhancenetworksafety.IDScanalsobeusedtomonitornetwork traffic (LI Yongzhong *et al*., 2008), thereby detecting if a system is being targeted by a network attack (S K Sharma *et al.*, 2012) such asaDoSattack.IDSsremaintheonlyproactive meansofdetectingandrespondingtothreatsthat stem from both inside and outside a corporate network.

Intrusion detection tools use several techniquestohelpthemdeterminewhatqualifies as an intrusion versus normal traffic (LI Yongzhong*etal*.,2008).Whetherasystemuses anomaly detection, misuse detection, target monitoring, or stealth probes, they generally fall into one of twocategories:

- Host-basedIDSs(HIDS)–examinedataheld on individual computers that serve as hosts. The network architecture of host-based (ReutersNewsService,2005)isagent-based, which means that a software agent resides on each of the hosts that will be governedby the system.

- Network-based IDSs (NIDS) – examine data exchanged between computers (Reuters News Service, 2005). More efficient host- basedintrusiondetectionsystemsarecapable ofmonitoringandcollectingsystemaudittrails in real time as well as on a scheduled basis, thus distributing both CPU utilization and networkoverheadandprovidingforaflexible means of securityadministration.

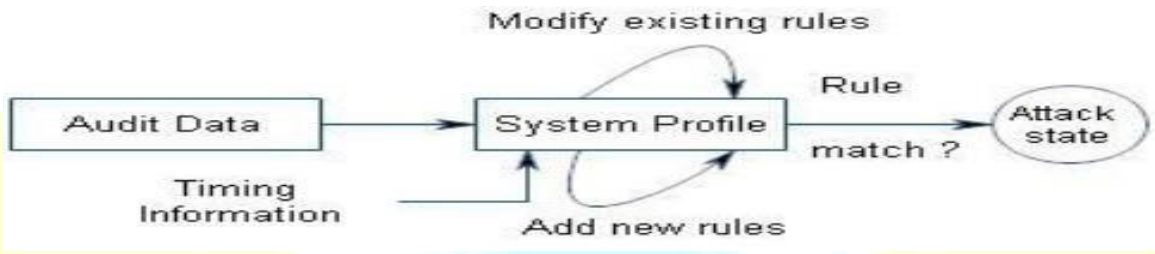
IDSscanalsobecategorizedaccordingtothe detection approaches they use (Levin (2000). Basically, there are two detection methods: misuse detection and anomaly detection.

The majordifference betweenthetwomethodsisthat misuse detection identifies intrusions based on features of known attacks while anomaly detection analyzes the properties of normal behavior. IDSs that employ both detection methodsarecalledhybriddetection-basedIDSs. Examples of hybrid detection-based IDSs are HybridNIDSusingRandomForestsandNIDES (D Dasgupta, 1998). The following subsections explainthetwodetectionapproaches.

MisuseDetection

Misuse detection catches intrusion in terms of the characteristics of known attacks. Anyaction

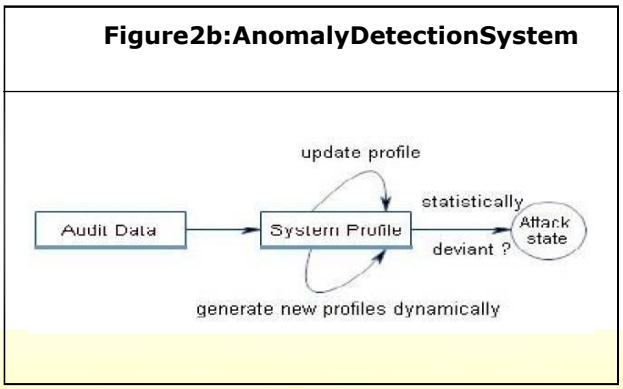
Figure2a:MisuseDetectionSystem



Thatconformstothepatternofaknownattackor vulnerabilityisconsideredasintrusive.Themain issues in misuse detection system are how to write a signature that encompasses all possible variationsofthepertinentattack.Andhowtowrite signatures that do not also match non-intrusive activity.BlockdiagramFigure2aofmisusebased detectionsystemisasfollowing.Misusedetection identifiesintrusionsbymatchingmonitoredevents to patterns or signatures of attacks. The attack Signaturesarethecharacteristicsassociatedwith successful known attacks The major advantage ofmisusedetectionisthatthemethodpossesses high accuracy in detecting known attacks. However, its detection ability is limited by the signature database. Unless new attacks are transformed into signatures and added to the database, misuse-based IDS cannot detect any attack of this type. Different techniques such as expert systems, signature analysis, and state transitionanalysisareutilizedinmisusedetection.

Anomaly DetectionSystem

It is based on the normal behavior of a subject (e.g., a user or a system). Any action that significantlydeviatesfromthenormalbehaviors considered as intrusive. That means if we could establish a normal activity profile for a system, then we can flag all system states varying from establishedprofile.Thereisaimportantdifference between anomaly based and misuse based technique that the anomaly based try to detect the compliment of bad behavior and misuse based detection system try to recognize the known bad behavior. In this case we have two



possibilities: (1) False positive: Anomalous activitiesthatarenotintrusivebutareflaggedas intrusive.(2)FalseNegative:Anomalousactivities thatareintrusivebutareflaggedasnonintrusive. TheblockdiagramFigure2bofanomalydetection system is asfollowing:

Anomaly detection assumes that intrusions areanomaliesthatnecessarilydifferfromnormal behavior.Basically,anomalydetectionestablishes a profile for normal operation and marks the activitiesthatdeviatesignificantlyfromtheprofile as attacks. The main advantage of anomaly detection is that it can detect unknown attacks (VaughnRandalandEvronGadi(2007);Zhenglie Li (2011). However, this advantage is paid for in terms of a high false positive rate because, in practice,anomalies are not necessarily intrusive. Moreover, anomaly detection cannot detect the attacks that do not obviously deviate from normal activities.As the number of new attacks increases rapidly, it is hard for a misuse detection approach to maintain a high detection rate. In addition, modeling attacks is a highly qualified and time- consuming job that leads to a heavy workload of maintaining the signature database.On the other hand, anomaly detection methods that discover the intrusions through heuristic learning are relatively easy to maintain.

When there is an intruder who has no idea of the legitimate user's activity patterns, the probability that the intruder's activity is detected as anomalous should be high. Four possibilities in such a situation, each with a non-zero probability.

- Intrusive but not anomalous: An IDS may fail to detect this type of activity since the activity is not anomalous.But,if theIDS detects such an activity, it may report it as a false negative because it falsely reports the absence of an intrusion when there is one.

- Not intrusive but anomalous: If the activity is not intrusive,but it is anomalous,an IDS may report it as intrusive. These are called false positives because an intrusion detection system falsely reports intrusions.
- Not intrusive and not anomalous: These are true negatives; the activity is not intrusive and should not be reported as intrusive.
- Intrusive and anomalous: These are true positives;the activity is intrusive

and much be reported assuch.

CONCLUSION

In this paper, we review IDS tools are becoming increasingly necessary. They round out the securityarsenal,workinginconjunctionwithother informationsecuritytools,suchasfirewalls,and allowforthecompletesupervisionofall network activity. It is very likely that IDS capabilities will become corecapabilities of network infrastructure (such as routers, bridges and switches) and operating systems. In future we would like to find out how data mining can help improve intrusion detection and most of all anomalydetection.Forthatpurposewehaveto understand how an IDS work to identify an intrusion.Byidentifyingboundsforvalidnetwork activity, data mining will aid an analyst to distinguishattackactivityfromcommoneveryday trafficonthenetwork.Thiswillrequire,Ibelieve, combination of multiple complicated methodsto coverallofthedifficultieswillmakeitevenmore timeconsuming.

REFERENCES

1. Alex Lam (2014), "New IPS to Boost Security,ReliabilityandPerformanceofthe Campus Network", Newsletter of Computing ServicesCenter.
2. Pfahringer B (2013), "Winning the KDD99 Classification Cup: Bagged Boosting", In SIGKDDExplorations.
3. Barbara D, Domeniconi C and Rogers J (2014), "Detecting Outliers Using Transduction And Statistical Testing", *Association for ComputingMachinery.*
4. Dasgupta D(2011),"AnArtificialImmuneSystem AsAMultiagentDecisionSupportSystem", *IEEEInternationalConferenceonSystems,Manand Cybernetics,* pp.3816-3820.
5. Reuters News Service (2005), FBI Agents bust 'Botmaster', November4.
6. Jelena Mirkovic, Sven Dietrich, David Dittrich and Peter Reiher (2005), *Internet*

Denial of Service: Attack and Defense Mechanisms, Prentice Hall PTR, ISBN 0131475738.

7. MaJandPerkinsS(2003),"OnlineNovelty Detection on Temporal Sequences", ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Washington,DC.

8. Levin (2000), "KDD-99 Classifier Learning Contest: LLSoft's Results Overview", SIGKDDExplorations.

9. LIYongzhong,YANGGeandXUJingZhao Bo (2008,) "A New Intrusion Detection Method Based on Fuzzy HMM", *IEEE*,Vol. 2, No.8.

10. Ihler A, Hutchins J and Smyth P (2006), "Adaptive Event Detection withTime-varying Poisson Processes", ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining(KDD),Philadelphia,PA.

11. Sharma S K, Pandey P and Tiwar S K (2012), "An Improved Network Intrusion Detection Technique Based On k-means ClusteringViaNaïveBayesClassification", *IEEE*, Vol. 2, No.2.

12. Tarem Ahmed, Boris Oreshkin and Mark Coates (2007), "Machine Learning ApproachestoNetworkAnomalyDetection", in Workshop on Tackling Computer Systems Problems with Machine Learning Techniques,McGillUniversityMontreal,QC, Canada.

13. Vaughn Randal and Evron Gadi (2007), "DNSAmplificationAttacks".

14. Zhenglie Li (2011) "Anomaly Intrusion Detection Method Based on K-Means Clustering Algorithm with Particle Swarm Optimization",Springer,Vol.4,No.2.