ISSN: 2277- 5528 Impact Factor- 4.015

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES& MANAGEMENT A DATA MINING BASED TECHNIQUE TO ERADICATE THE CHALLENGES BEING FACED WITH THE TRADITIONAL IDS - SIMULATION & RESULT

Mohammed Imroz Malik and Kamal Niwaria

Department of Electronics & Communication Engineering, SRK University, Bhopal (MP) India

e-mail: imroz.19@gmail.com, kamalniwaria@gmail.com,

*Corresponding Author: imroz.19@gmail.com

ABSTRACT

The article is going to be present a general idea for intrusion detection system which will enhance efficiency as compare existing intrusion detection system. The proposed concept is using data mining techniques. In this clustering and classification data mining technique has applied for anomaly detection field of intrusion detection. Anomaly learning approaches are able to detect attacks with high accuracy and to achieve high detection rates. However, the rate of false alarm using anomaly approach is equally high. In order to maintain the high accuracy and detection rate while at the same time to lower down the false alarm rate, the proposed technique is the combination of three

INTRODUCTION

This paper is going to be present general idea on a new proposed concept for intrusion detection system which will enhance efficiency as compare existing intrusion detection system. The proposed concept is using data mining techniques. Data mining techniques have been successfully applied in many different fields including marketing, manufacturing, process control, fraud detection, and network management. Over the past five years, a growing number of research techniques have applied data mining to various problems in intrusion detection. In this will apply to data mining for anomaly detection field of intrusion detection. Presently, it is unfeasible for several computer systems to affirm security to network intrusions with computers increasingly getting connected to public accessible networks (e.g., the Internet). In view of the fact that there is no ideal solution to avoid intrusions from event, it is very significant to detect them at the initial moment of happening and take necessary actions for reducing the likely damage. One approach to handle suspicious behaviors inside a network is an intrusion detection system (IDS). For intrusion detection, a wide variety of techniques have been applied specifically, data mining techniques, artificial intelligence technique and soft computing techniques. Most of the data mining techniques like association rule mining, clustering and classification have been applied on intrusion detection, where classification and pattern mining is an important technique.

DATA MINING

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Although data mining is a relatively new term, the technology is not. Companies have used powerful computers to sift through volumes of supermarket scanner data and analyze market research reports for years. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost.

ISSN: 2277- 5528 Impact Factor- 4.015

DATA MINING WORK

Data mining is primarily used today by companies with a strong consumer focus - retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff, skills, and "external" factors such as economic indicators, competition, and customer demographics. And, it enables them to determine the impact on sales, customer satisfaction, and corporate profits. Finally, it enables them to "drill down" into summary information to view detail transactional data. With data mining, a retailer could use point-of-sale records of customer purchases to send targeted promotions based on an individual's purchase history. By mining demographic data from comment or warranty cards, the retailer could develop products and promotions to appeal to specific customer segments. While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. Generally, any of four types of relationships are sought:

- Classes: Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.
- Clusters: Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.
- Associations: Data can be mined to identify associations. The beer-diaper example is an example of associative mining.
- Sequential patterns: Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

DATA MINING CONSISTS OF FIVE MAJOR ELEMENTS

- Extract, transform, and load transaction data onto the data warehouse system.
- Store and manage the data in a multidimensional database system.
- Provide data access to business analysts and information technology professionals.
- Analyze the data by application software.
- Present the data in a useful format, such as a graph or table.

ANALYSIS

- Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- Genetic algorithms: Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.
- Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square

Automatic Interaction Detection (CHAID). CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID.

- Nearest neighbor method: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k 1). Sometimes called the k-nearest neighbor technique.
- Rule induction: The extraction of useful if-then rules from data based on statistical significance.
- Data visualization: The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationship

ARCHITECTURE



Figure 4.1: Block Diagram

Here we outline a data mining approaches for designing intrusion detection models. The Basic idea behind this is that apply various data mining technique in single to audit data to compute intrusion detection models, as per the observation of the behavior in the data. In the proposed work are the combining three most useable data mining techniques into single concept and presenting architecture shown in figure 4.2. In proposed technique, use Hierarchical clustering, C5.0 algorithm and CHAID approach. First apply the hierarchical algorithm to the given dataset to split the data records into normal cluster and anomalous clusters. It specifies the number of clusters as five to the hierarchical and clusters the records in the dataset into normal cluster and anomalous clusters. The anomalous clusters are U2R, R2L, PROBE, and DoS. The records are labeled with the cluster indices. Then, divide the data set into two parts. One part is used for

ISSN: 2277- 5528 Impact Factor- 4.015

training and the other one is used for evaluation. In training phase, apply the labeled records to the C5.0 for training purpose. The C5.0 classifier is trained with the labeled records. Then, apply the rest of unlabeled records to the C5.0 for classification. The C5.0 classifier will classify the unlabelled record into normal and anomalous clusters. Finally apply CHAID which is also the classifier that is doing exact match of each attribute values all to gather and thus removes the strong independence assumption. The Proposed work consists of clustering, classification where proposed architecture as shown in figure 4.2.



Figure 4.2: Architecture of the IDS

CONCEPT

Here proposed concept are going to be present general idea as showing in figure 4.1 for intrusion detection system which will enhance efficiency as compare existing intrusion detection system. The proposed concept is using data mining techniques. In this clustering and classification data mining technique has applied for anomaly detection field of intrusion detection. Anomaly learning approaches are able to detect attacks with high accuracy and to achieve high detection rates. However, the rate of false alarm using anomaly approach is equally high. In order to maintain the high accuracy and detection rate while at the same time to lower down the false alarm rate, the proposed technique is the combination of three learning techniques. For the first stage in the proposed technique, this grouped similar data instances based on their behaviors by utilizing a hierarchical clustering as a pre-classification component. Next, using C5.0 classifier this classified the resulting clusters into attack classes as a final classification task. This found that data that has been misclassified during the earlier stage may be correctly classified in the subsequent classification stage. At last CHAID classification is applied. Following is the proposed IDS which divided into following module:

1. Database Creation (Suggested Technique)

•Download and Rearranged KDD 99'

•Data Formation and Re-Processing of KDD 99' (Training and Testing Data Set Prepration)

- Data mining Techniques • Cluster Technique
- Cluster Technique
- o Hierarchical Clustering
- Classification
- o C5.0
- o CHAID
- 3. Proposed System

2.

•K-Mean Clustering

•K-Mean Clustering with Naïve Bayse classification

•K-Mean with Naïve Bayse classification and Decision Table Majority Rule Based Approach

- Hierarchical Clustering
- Hierarchical Clustering with C5.0 classification
- Hierarchical Clustering with C5.0 classification and CHAID Classification
- 4. Performance

•Time Analysis

•Memory Analysis •CUP Analysis

ALGORITHM

Input: Dataset KDD, a sample K, Normal Cluster NC, Abnormal cluster AC, c is the number of clusters and d is the distance between them, ch1,ch2,ch3,ch4,ch5 are Nodes i1,i2,i3,i4 are the category Output: K is abnormal or normal

Algorithm Hybrid

A) First apply Hierarchical clustering

1) Firstly load data into a root cluster and we start with one cluster and successively split clusters to produce others, more and more samples are clustered together in a hierarchical manner.

2) For Every data point:

3) Find out the distance from the data point to every cluster.

<u>Begin</u>

<u>Initialize</u> c; c' = n; $Di = \{xi\}$; i = 1,...,n**Do**

c' = c' - 1

4) Find nearest clusters Di and Dj

5) Merge Di and Dj <u>Until</u> c = c'<u>Return</u> c clusters <u>End</u>

6) To find the nearest clusters in step 4, the following clustering criterion function is used:

dmin(Di, Dj) = min || x - x' ||, where $x \in Di$ and $x' \in Dj$

7) The merging of the two clusters in step 6 simply corresponds to adding an edge between the nearest pair of nodes in Di and Dj. Also, if instead of terminating after a predetermined number of clusters have been obtained; it is possible to set the termination criteria to stop when the distance between nearest clusters exceeds a predetermined threshold.

B) Apply C5.0 Classification

1) For each Clusters C in KKDi in test data do

If C is i1

Ch1=c Else

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & MANAGEMENT

If C is i2 Ch2=c Else If C is i3 Ch3=c Else If C is i4 Ch4=c Else Ch5=c **until** end of data set

2) Collect data from dataset in the form of Normal/Abnormal and apply those data to the CHAID Decision Table Majority rule based approach and build condition for the action like training/testing normal data set D.

C) CHAID

3) Preparing predictors. The first step is to create categorical predictors out of any continuous predictors by dividing the respective continuous distributions into a number of categories with an approximately equal number of observations. For categorical predictors, the categories (classes) are "naturally" defined.

If (c is not equal to ch1,ch2,ch3,ch4) Then c is Normal Otherwise c is abnormal

SIMULATION AND RESULTS

This IDS system is implemented in GUI is used as a tool for database creation and management. which can be seen as shown in fig



Fig 5.1 First Form "Login Form"

Computer	Mozilla Firefox	1			
Network	Net Protector 2014	6			
Recycle Bin	NetBeans IDE	DIISuite	IDS System		1
uTorrent.	Reliance 3G	SpeedAna	 <u>k-Means clustering</u> k-Means Naïve Bayes and Decision Tree 	 k-Means and Naïve Bayes Result 	Go
Adobe Reader 9	VLC media player	MS TurboC+	Hierarchical clustering and C5 9 algorithm		
Huawei Connecti	You J Tube Xilisoft Downlo	XAMPP Contr	Hierarchical clustering C5.0 algorithm and CHAID Result		Go
Mobile Partner	MMX352G 3G USB	Mobile Broadband			Windows 7 Build 7601
1			👋 🜔 🔬 🖭		Desktop * - 🎲 🏴 🔐 3:59 PM 01/10/2013

Fig 5.2 Second Form "Selection of Algorithms"

¥ Cut	<u>è</u>				К	Mean Result									AA F
Ца Сору	DOS		R2L			U	2R		P	rob			Norm	2A	ab R
IFormat Pai	Protocol Attack_T		Protocol_Type	Attack_Type		Protocol_Type	Attack_Type	Prot	ocol_Type	Attack_Type		Protocol_Type	Attack_Type	styles *	là s
Clipboard	tcp neptune.		tcp	ftp_write.	-	tcp	perl.	tcp		portsweep.		tcp	normal.	0 83	E
	tcp neptune.	-	tcp	imap.	-	tcp	buffer_overflow.	tcp		portsweep.	0	tcp	normal.		
	icmp smurf.		tcp	ftp_write.		tcp	loadmodule.	tcp		portsweep.		tcp	normal.		
	icmp smurf.		tcp	multihop.		tcp	perl.	tcp		portsweep.		tcp	normal.		
	icmp smurf.		tcp	ftp_write.		tcp	rootkit.	tcp		portsweep.		tcp	normal.		
	icmp smurf.		tcp	ftp_write.		tcp	rootkit.	tcp		portsweep.		tcp	normal.		
	icmp smurf.	-	tcp	imap.	τ.	tcp	rootkit.	tcp		ipsweep.	-	tcp	normal. +		
	885		1015			145		4154				78932			
	Start 4:32:47:705		End 4:32:48:704			CPU Uses	84%	Tota	Record	85131					
												Loa	ad Data	1000	
	-										-				
	protocol_type				ftn	ning_attack_types				-		Ар	ply Algo		
	μαφ				imar).					2	Q			
					ftp	write.						-			
	tcp				mult	ihop.									
	tcp				perl										
	tcp				buff	er_overflow.									
	tcp				load	module.									
	tcp				perl										
	tcp				root	kit.									
	tcp				root	kit.									
	tcp				root	kit.									
	ten				1001	NL.									
	tro				peri	or overflow									
	0.p				Dun	er_overnow.									
	tro				norr	nal									
	tcp				norr	nal.					-				
	tcp tcp										_				
	tcp tcp tcp	_													
	tcp tcp tcp														
	tcp tcp tcp			fto	write.								Editing		
	tcp tcp tcp			fo	write.		10						Editing		

Fig 5.3 Execution of "K-Mean Approach"

Hierarchical C5.0 and CHAID pe fraing_attack_types		Decument? Microsof	\$1Mard				- 0
Herarchical C5.0 and CHAD pe raining_attad_types ftp_write. inditiop. ftp_write. inditiop. perf. Apply Algo codt. perf. codt. codt. contal. codt. contal. contal. contal. co	2						
pe traing_sttod_types fb_write. fb_write.		Hierarchical C5.0 and CHAID					
fb_write. Ioda Data fb_write. Ioda Data fb_write. Ioda Data pel. Ioda Data pel. Ioda Data pel. Ioda Data posteriow. Ioda Data pel. Ioda Data posteriow. Ioda Data perd. Ioda Data posteriow. Ioda Data perd. Ioda Data	protocol_type	training_attack_types					
insp. insp. insp. 1 10/100p. 10/100p. 10/100p. 1 10/100p. 10/100p. 10/100p. 10/100p. 1 10/100p. 10/100p. 10/100p. 10/100p. 10/100p. 1 10/100p.	tcp	ftp_write.				Lord Date	
fp_write. nullbop. perl. loadmodule. perl. rootit. rootit. rootit. rootit. perl. rootit. rootit. perl. rootit. perl. rootit. perl. normal. norm	tcp	imap.				Load Data	
multings. pert. buffer_overflow. pert. port.	tcp	ftp_write.					
perl, Apply Ago bodmod/e, Bodmod/e, perl, Incolut. rootist. Incolut. perl, Incolut. normal. Incormal. <td>φ</td> <td>multhop.</td> <td></td> <td></td> <td></td> <td></td> <td></td>	φ	multhop.					
buffe_overflow. loadinoalie. perf. rootht. rootht. <tr< td=""><td>φ</td><td>perl.</td><td></td><td></td><td></td><td>Apply Algo</td><td></td></tr<>	φ	perl.				Apply Algo	
loadmodile. operl. rootht. rootht. rootht. rootht. rootht. rootht. rootht. rootht. perl. buffe_yverflow. normal.	cp	buffer_overflow.					
peri. Image: Second Secon	φ	loadmodule.					
Incodel. Herarchical CS.0 and CHAD nooblt. nooblt. nooblt. DOS nooblt. DOS nooblt. DOS nooblt. DOS obdfe_oreflow. RAL normal. RAL <	φ	peri.					
Inotel: Hierarchical C5.0 and CHAD inotel: inotel: inotel:	φ.	rootkit.					
notbit: notbit: perl. perl. buffe_overflow. perl. normal. perl.		rootkit.			Hierarch	cal C5.0 and CH	AID
notifit: DOS 885 Total Recore 85131 buffer_overflow. normal. normal. <td< td=""><td>top</td><td>rootkit.</td><td></td><td></td><td></td><td></td><td></td></td<>	top	rootkit.					
peri- porti- porti- </td <td>0</td> <td>rootkit.</td> <td></td> <td></td> <td>-</td> <td></td> <td></td>	0	rootkit.			-		
buffer_overflow. normal. normal. normal.	THE STREET	perl.		DOS	885	Total Recore	85131
normal. normal. normal. normal. normal. normal. normal. normal. normal. 145 normal. normal. normal. 145 normal. 145		buffer overflow.					
normal. R2L 1015 CPU Uses 60% normal. normal. 1015 CPU Uses 60%		normal					
normal. normal.	Ф Ф	normal					
Informal. Informal.	φ	normal		R2L	1015	CPU Uses	60%
normal. U2R 145 normal. normal. 145 normal. normal. 145 normal. 145 145	φ 	normal.					
normal. U2R 145 normal. normal. 145 normal. normal. 145 normal. 145 145 normal. 145 145 normal. 145 145 normal. 145 145	φ	normal					
Initial Initial normal. normal. normal. normal. normal. 1 normal. 1 normal. 1 normal. 1 normal. 1	φ	normal.		U2R	145		
normal. normal. normal. normal. normal. normal. normal. normal.	φ	normal.					
normal. Prob 4154 Start 4/324 Prob 4154 Start 4/324 Prob 4154 Start 4/324 Prob 4154 Start 4/324 Prob 4154 Start	φ 	normal.					
pormal. pormal. pormal. pormal.	φ	normai.				1	
pormal.	φ	normai.		Prob	4154	Start	4:43:24:311
nomal,	φ	normai.					
normal.	φ	normal.					
79000 End 4/200E	ф	normal.			70022	End	4-42-25-724
normal. Norm 10952 End 4.45.25	φ	normal.		Norm	10932	Ellu	4.45.25.751
normal.	φ	normal.					
normal.	.cp	normal.		-			
normal. normal. pormal. normal.	tep tep tep	normal. normal. normal.	v	Norm	78932		End
	t at 1 Martin 8 M			_	mundum	1005	
	ge: 1 of 1 Words: 8 🔇					100% -	
Words: 8 🗭	9 📋 🖸 🏉 🦉					- 🚜 P* 🕻	4:44 PM 01/10/201

Fig 5.4 Execution of "K-Mean and Naïve Byes" Approach

	training_attack_types				
tcp	Iftp write,			Load Data	
top	imap.				
tcp	ftp write,	 			
tcp	multihop.			Apply Algo	
tcp	peri.				
tcp	buffer_overflow.				
tcp	loadmodule.				
tcp	peri.				
tcp	rootkit.	k-N	leans K-ne	arest and Decis	ion Tree
tcp	rootkit.	R-I	is and it in	arestand Decis	
tcp	rootkit.				
top	rootkit.	DOS	oor	Total Decore	05121
top	perl.	003	005	Total Recore	05151
tcp	buffer_overflow.				
tcp	normal.	0.01	1015	CDUUIses	710/
tcp	normal.	RZL	1015	CPU Uses	/1%
tcp	normal.				
tcp	normal.	1120	145		
tcp	normal.	Van	145		
tcp	normal.				
tcp	normal.	Droh	4154	Start	4-21-52-256
tcp	normal.		1131	Jun	1.52.52.550
tcp	normal.				
tcp	normal.	Norm	79032	End	4:21:52:040
tcp	normal.		70952	Linu	4.51.52.545
tcp	normal.				
tcp	normal.				
tcp	normal.				
tcp	normal.				
(CD)	inormal				

Fig 5.4 Execution of "K-Mean and Naïve Byes and Decision Tree" Approach

Computer	Mozilla Firefox	1										
Network	Net Protector 2014	6										
Recycle Bin	NetBeans IDE 7.0.1	DilSuite				\$	Start Time	End Time	CPU Uses			
Ju Torrent	Reliance 3G	SpeedAnal				K-Mean						
Adobe Reader 9	VLC media player	MS TurboC++				k-Means Naive Bayes & Decision Tree						
Awei Huawei Connecti	You Tube Tube Xilisoft Downlo	XAMPP Contr						Go				
Mobile Partner	MMX352G 3G USB	Mobile Broadband									This serves of Mile day	Windows 7 Build 7601
@			1	\bigcirc	100 m					Desktop **	🔺 🍕 🏴 🛱	4:02 PM 01/10/2013

Fig 5.5 Result of Above Three Approach

Computer	Mozilla Firefox	1								
Network	Net Protector 2014	r 6								
				<u>ه</u>						
Recycle Bin	NetBeans IDE 7.0.1	DIISuite			Start Time	End Time	CPU Uses			
u Torrent	Reliance 3G	SpeedAnal		K-Mean	4:32:47:705	4:32:48:704	84%			
profient	Renalice 30	SpecoAllal		k-Means & Naïve Bayes	4:32:21:887	4:32:22:948	74%			
Adobe Reader 9	VLC media player	MS FD TurboC++		k-Means Naïve Bayes & Decision Tree	4:31:52:356	4:31:52:949	71%			
	You Tube	83				Go				
Huawei Connecti	Xilisoft Downlo	XAMPP Contr								
Mobile Partner	MMX352G 3G USB	Mobile Broadband							This convol Windows	Windows 7 Build 7601
			(3)				<u> </u>	Desktop		4:33 PM

Fig 5. 7 Result of Above Three Approach

ISSN: 2277- 5528 Impact Factor- 4.015

		<u>*</u>					P
Computer	Mozilla			Hierarchical Result			
	FIFETOX	DOS	R2L	U2R	Prob	Norm	I C D
		Protocol Attack_Ty	Protocol_Type Attack_Type	Protocol_Ty Attack_Type	Protocol_Ty Attack_Type	Protocol_Type Attack_Type	
		tcp neptune.	tcp ftp_write.	🔺 tcp perl. 🚽	tcp portsweep.	tcp normal.	
		tcp neptune.	tcp imap.	tcp buffer_overflo	tcp portsweep.	tcp normal.	
Network	Net Protect	icmp smurf.	tcp ftp_write.	tcp loadmodule.	tcp portsweep.	tcp normal.	
	2014	icmp smurf.	tcp multihop.	tcp perl.	tcp portsweep.	tcp normal.	
	-	icmp smurf.	tcp ttp_write.	tcp rootkit.	tcp portsweep.	tcp normal.	
		icmp smurf.	tcp ttp_write.	tcp rootkit	tcp portsweep.	tcp normal.	
F		icinp sinuii	icp inap.		lich ibsweep.	r icp normai. 🔫	
Recycle Bin	NetBeans II	885	1015	145	4154	78932	
	7.0.1	Stat 44042-044	5-4 44045400	70%	Tetal Descent OF 424		
		Start 4.40.13.944	End 4.40.15.192	CPU Uses 79%	Total Record 85131	Load Data	
	@						
		protocol_type		training_attack_types			
	Deliana 20	tcp		ftp_write.	*	Apply Algo	
proment	Kellance 34	tcp		imap.			
		tcp		rtp_write.			
		tep		naidrop.			
8	_	trp		buffer overflow			
		trp		loadmodule			
		top		perl.			
Adobe Peader 0	VLC media	tcp		rootkit.			
Reader 9	piayer	tcp		rootkit.			
		tcp		rootkit.			
	You	tcp		rootkit.			
	Tube	tcp		perl.			
AWE		tcp		buffer_overflow.			
Huawei	Ailisoft	tcp		normal.			
connecti	bownio	tcp		normal.			
		tcp		Inormai.	-		
		Mark 1					Window
Mobile	MMX352G	Mobile					Puild 7
Partner	30 O2B	Broadband				This con	Build /
	-					»	4.40 DIA
1		o 😂 🔮 🚺				Desktop 🛥 🚓	01/10/2013

Fig 5.7 Execution of "Hierarchical Clustering" Approach

B	Home Insert	Page Layout	Reference	es Ma	ilings	Review 1	View	Document3 - N	licrosoft Word					-	
ß						-		Τ.							nd
Paste						Hier	archical and C5.0								elect
	protocol_type						training_attack_t	types							ting
	tcp						ftp_write.			-			Load Data		
	tcp						imap.			<u></u>			Luau Data		
	tcp						ftp_write.								
	tcp						multihop.						Apply Algo		
	tcp						perl.						Apply Algo		
	tcp						buffer_overflow.								
	tcp						loadmodule.								
1	tcp						perl.					Hie	rarchical and C5.0		
	tcp						rootkit.								
	tcp						rootst.								
	tcp						rootkit.								
	tcp						roobst.				DOS	885	Total Recore	85131	
	tro						buffer overflow								
	trp						pormal								
	tcp						normal.					4045	00000000	770/	
	tro						normal				R2L	1015	CPU Uses	77%	
	tcn						normal.								
	tcp						normal.								
	tcp						normal.				U2R	145			
	tcp						normal.								
	tcp						normal.								
	tcp						normal.				Deeb	4454	Chard	4.07.44.400	
	tcp						normal.				Prob	4154	Start	4.37.41.422	
	tcp						normal.								
	tcp						normal.						-	_	-
	tcp						normal.				Norm	78932	End	4:37:42:670	
	tcp						normal.								-
	tcp						normal.				_				
	tro						normal			-					
															_
Page 1 of	1 Worder A	6										10		60%	- m
rage: 1 01	1 Wolds: 4			-							_				
1			1	\bigcirc										🚜 🍽 🕼	4:39 PM 01/10/2013

Fig 5.8 Execution of "Hierarchical Clustering and C5.0" Approach

	Hierarchical C5.0 and CHAID					
rotocol_type	training_attack_types					
0	ftp_write.	×				
	imap.			Load	Jata	
š	ftp_write.					
	multihop.		C			
1	perl.			Apply	Algo	
	buffer_overflow.					
	loadmodule.					
	perl.					
	rootkit.					
	rootkit.			Hierarchical C5	0 and CH	AID
	rootkit.					
	rootkit.					
	perl.		DOS	885 Tota	Recore	85131
	buffer_overflow.					
	normal.					
	normal.			1015 CPU	leas	60%
	normal.		R2L	1015 000	0363	00%
	normal.					
	normal.					
	normal.		U2R	145		
	normal.					
	normal.					
	normal.		Droh	4154 Start		4-43-24-3
	normal.		1100	4134		4.45.24.5
	normal.					
	normal.					1
	normal.		Norm	78932 End		4:43:25:73
	normal.					
	normal.					
	normal	•				

Fig 5.9 Execution of "Hierarchical Clustering, C5.0 and CHAID" Approach



Fig 5.10 Results of Above Three Approach

Computer	Mozilla Firefox	1							
Network	Net Protector 2014	6							
Recycle Bin	NetBeans IDE 7.0.1	DilSuite		♣ Hierarchical	Start Time 4:40:13:944	End Time 4:40:15:192	CPU Uses		
µTorrent	Reliance 3G	SpeedAnal		Hierarchical and CS.V algorithm	4:37:41:422	4/37/42/670	(1%)		
Adobe Reader 9	VLC media player	MS TurboC++		Hierarchical C5.0 algorithm and CHAID	4:43:24:311	4:43:25:731	60%		
Awer Huawei Connecti	You Tube Zilisoft Downlo	XAMPP Contr			(Go]		
Mobile Partner	MMX352G 3G USB	Mobile Broadband						This copy of Window	Windows 7 Build 7601 vs is not genuine
0			1					Desktop 🎽 🔺 🏴 😭	4:47 PM 01/10/2013

Fig 5.11 Results of Above Three Approach

CONCLUSION

As information systems have become more comprehensive and a higher value asset of organizations, intrusion detection systems has been incorporated as elements of operating systems, although not typically applications. Intrusion detection involves determining that some entity, an intruder, has attempted to gain, or worse, has gained unauthorized access to the system. This research shows that benchmarking intrusion detections systems can be done effectively. In this work design and develop more advanced data mining techniques, it will be very hard to evaluated proposed intrusion detection systems. The amount of customization of data mining techniques that goes into effectively using one, as well as the ever-changing number of viable network exploits makes it impossible at this time. The speed of operation of During Data mining technique is faster. During testing, proposed IDS run in few second to get output. In this no debugging is required. This is due to the high amount of string optimization involved through data mining technique.

REFERENCES

[1] Eric Bloedorn, Alan D. Christiansen, William Hill "Data Mining for Network Intrusion Detection: How to Get Started" 2001.

[2] Skorupka, C., J. Tivel, L. Talbot, D. Debarr, W. Hill, E. Bloedorn, and A. Christiansen 2001. "Surf the Flood: Reducing High-Volume Intrusion Detection Data by Automated Record Aggregation," Proceedings of the SANS 2001 Technical Conference, Baltimore, MD

[3] Sumathi, S.; Sivanandam, S. N.: Introduction to Data Mining and its Applications. Springer, 2006.

[4]

ISSN: 2277- 5528 **Impact Factor- 4.015**

Fayyad, Piatetsky-Shapiro, Smyth: From Data Mining to Knowledge Discovery in Databases. AI MAgazine, 1996.

Roiger, Richard J.; Geatz, Michael W.: Data Mining: A Tutorial- Based Primer. Addison Wesley, [5] 2003

[6] MIT linconin labs, 1999 ACM Conference on Knowledge Discovery and Data Mining (KDD) Cup dataset, http://www.acm.org/sigs/sigkdd/kddcup/index.php?section=1999

The KDD Archive. 1999. KDD99 cup dataset. [7]

http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

M. Tavlle, E. Bagheri, W. Lu, and A. A. Gorbani, "A detailed analysis of the KDD CUP 99 Data [8] Set," Proc. of IEEE Symposium Computational Intelligence for Security and Defense Applications (CISDA'09), pp. 1-6, 2009.

James P. Anderson, "Computer security threat monitoring and surveillance," Technical Report 98-[9] 17, James P. Anderson Co., Fort Washington, Pennsylvania, USA, April 1980.

[10] D. E. Denning, "An intrusion detection model," IEEE Transaction on Software Engineering, SE-13(2), 1987, pp. 222-232.

LI Min "Application of Data Mining Techniques in Intrusion Detection" 2005 [11]

[12] KDD. (1999). Available at http://kdd.ics.uci.edu/databases/ - kddcup99/kddcup99.html

J. Han, and M. Kamber. Data mining: Concepts and techniques. Morgan Kaufmann. San Francisco, [13] CA. 2001.

[14] Hansen, Hans Robert; Neumann, Gustaf: Wirtschaftsinformatik I. Lucius & Lucius, 2001.

[15] Hipp, Jochen; Guentzer, Ulrich; Nakhaeizadeh, Gholamreza: Algorithms for Association Rule Mining - A General Survey and

http://www.seattlerobotics.org/encoder/mar98/fuz/flindex.html [16]

- [17] Northcutt and Novak, Network Intrusion Detection (3rd edition), New Riders, 2003.
- Barbará and Jajodia, Applications of Data Mining in Computer Security, Kluwer, 2002. [18]
- [19] Han and Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2001.
- Hand, Mannila and Smyth, Principles of Data Mining, MIT Press, 2001. [20]