Cluster Analysis of Smurf Type of Denial of Service Attack

Jigang Zheng[1, a*] and Jingmei Zhang[2, b]

[1]Department of Mathmatic, Baoshan College, Baoshan, Yunnan, 678000, China.

[2]Library of Baoshan College, Baoshan, Yunnan, 678000, China.

[a]6913641@qq.com, [b]279619568@qq.com

Keywords: Intrusion detection; Characteristic property; Weka; Cluster analysis

Abstract. Smurf denial of service attacks in the more common types of attacks, extracts the KDDCUP data set of denial of service the Smurf attack types, using the Weka to the attribute characteristic and cluster analysis, according to the similarity of attributes divided into 4 classes, and analyses the characteristics of various types, help to understand the network intrusion detection data set of internal rules.

Introduction

Denial of service (DoS) attack is currently widely used by hackers as a means of attack, mainly divided into SYN, Flood Smurf and Fraggle and other 3 kinds of [1].Refers to the so-called Smurf attack, attackers on the remote machine to send an ICMP response service request and the target host is a host IP address, but the broadcast address for a network the request packet source IP not initiated the attack of the IP address, but disguised to attack the host IP address [2].

Determine Data Mining Objectives

The experimental data set is derived from the "KDDCUP.data_10_percent" subset of the KDDCUP99 data set [3].Denial of service attacks accounted for more than data sets, Smurf attack types accounted for denial of service attacks total 71.73, visible Smurf attack is DoS attack is a common one, it is necessary to study it, and reduce the harm to DOS.

The data set each record contains the first 41 fixed feature attributes and the last 1 attack types identified, Smurf attack records have 32 attribute values are fixed and removed. That duration is 0,protocol_type is icmp, service is ecr_i, flag is SF, dst_bytes is 0,land is 0,wrong_fragment is 0,urgent is 0,hot is 0,num_failed_logins is 0,logged_in is 0,num_compromised is 0,root_shell is 0,su_attempted is 0,num_root is 0,num_file_creations is 0,num_shells is 0,num_access_files is 0,num_outbound_cmds is 0,is_host_login is 0,is_guest_login is 0,count is 0, serror_rate is 0,srv_serror_rate is 0,rerror_rate is 0,srv_rerror_rate is 0, same_srv_rate is 1, diff_srv_rate is 0,srv_diff_host_rate is 0, dst_host_srv_diff_host_rate is 0, dst_host_srv_serror_rate is 0,dst_host_srv_rerror_rate is 0.Left 10 feature attributes, ARFF format visualization results shown in Figure 1,through the Visualize all Preprocess visual interface, you can see the data classification summary visualization map, as shown in Fig. 2.

No.	src_bytes Nominal	srv_count Numeric	dst_host_count Numeric	dst_host_srv_count Numeric	dst_host_same_srv_rate Numeric	dst_host_diff_srv_rate Numeric	dst_host_same_src_port_rate Numeric	dst_host_serror_rate Numeric	dst_host_rerror_rate Numeric	class Nominal
1	1032	316.0	148.0	3.0	0.02	0.02	0.02	0.0	0.0	smurf.
2	1032	511.0	158.0	13.0	0.08	0.02	0.08	0.0	0.0	smurf.
3	1032	511.0	168.0	23.0	0.14	0.02	0.14	0.0	0.0	smurf.
4	1032	510.0	178.0	33.0	0.19	0.02	0.19	0.0	0.0	smurf.
5	1032	509.0	188.0	43.0	0.23	0.02	0.23	0.0	0.0	smurf.
6	1032	511.0	198.0	53.0	0.27	0.02	0.27	0.0	0.0	smurf.
7	1032	511.0	208.0	63.0	0.3	0.01	0.3	0.0	0.0	smurf.
8	1032	509.0	218.0	73.0	0.33	0.01	0.33	0.0	0.0	smurf.
9	1032	511.0	228.0	83.0	0.36	0.01	0.36	0.0	0.0	smurf.
10	1032	511.0	238.0	93.0	0.39	0.01	0.39	0.0	0.0	smurf.
11	1032	511.0	248.0	103.0	0.42	0.01	0.42	0.0	0.0	smurf.
12	1032	511.0	255.0	113.0	0.44	0.01	0.44	0.0	0.0	smurf.
13	1032	511.0	255.0	123.0	0.48	0.01	0.48	0.0	0.0	smurf.
14	1032	511.0	255.0	133.0	0.52	0.01	0.52	0.0	0.0	smurf.
15	1032	511.0	255.0	143.0	0.56	0.01	0.56	0.0	0.0	smurf.
16	1032	511.0	255.0	153.0	0.6	0.01	0.6	0.0	0.0	smurf.
17	1032	511.0	255.0	163.0	0.64	0.01	0.64	0.0	0.0	smurf.
18	1032	511.0	255.0	173.0	0.68	0.01	0.68	0.0	0.0	smurf.
19	1032	511.0	255.0	183.0	0.72	0.01	0.72	0.0	0.0	smurf.
20	1032	511.0	255.0	193.0	0.76	0.01	0.76	0.0	0.0	smurf.
21	1032	511.0	255.0	203.0	0.8	0.01	0.8	0.0	0.0	smurf.
22	1032	511.0	255.0	213.0	0.84	0.01	0.84	0.0	0.0	smurf.
23	1032	511.0	255.0	223.0	0.87	0.01	0.87	0.0	0.0	smurf.
24	1032	511.0	255.0	233.0	0.91	0.01	0.91	0.0	0.0	smurf.
25	1032	511.0	255.0	243.0	0.95	0.01	0.95	0.0	0.0	smurf.
26	1032	511.0	255.0	253.0	0.99	0.01	0.99	0.0	0.0	smurf.
27	1032	511.0	255.0	255.0	1.0	0.0	1.0	0.0	0.0	smurf.
28	1032	511.0	255.0	255.0	1.0	0.0	1.0	0.0	0.0	smurf.

Figure 1. ARFF format visualization

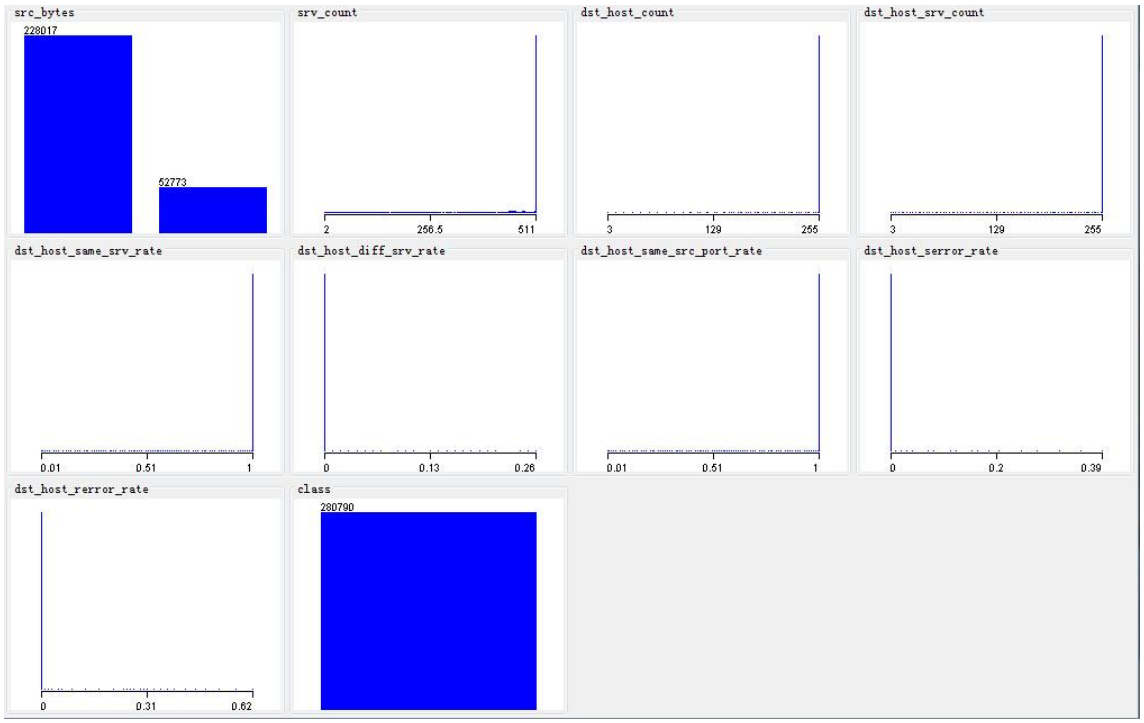

Figure 2. Attribute visualization

Cluster Analysis

Weka is one of the most complete data mining tools, and is recognized as one of the most famous open source projects in the open source project [4]. Provide clope, cobweb, DBSCAN,EM, Farthest First, Filtered Clusterer, Make Density Based Cluster, optics, SIB, SimpleK Means, XMeans a total of 11 kinds of clustering algorithm, in the experiment, this paper analysis based on, algorithm Filtered Clusterer clustering of experimental data. Filtered Clusterer algorithm is an integrated approach, which can be used to filter the data of any filter,and can be used in any way to cluster[5].

Set the parameters for "-N -A 5 weka. core. Euclidean Distance '-R first-last weka.clusterers.SimpleK Means' -I 500 -S 10", the results of the cluster mining as shown in Table 1.

Table 1 Results of cluster analysis

Attribute	Full Data (280790)	Clustered 0 (193082)	Clustered 1 (393)	Clustered 2 (17834)	Clustered 3 (34542)	Clustered 4 (34939)
src_bytes	1032	1032	1032	520	1032	520
srv_count	507.0331	511	241.4707	459.3776	509.2419	510.2393
dst_host_count	254.9815	255	250.1883	254.8148	255	255
dst_host_srv_count	254.9081	255	201.8753	254.8148	254.9527	255
dst_host_same_srv_rate	0.9997	1	0.7969	1	0.9998	1
dst_host_diff_srv_rate	0	0	0.0214	0	0	0
dst_host_same_src_port_rate	0.9997	1	0.7969	1	0.9998	1
dst_host_serror_rate	0	0	0.0079	0	0	0
dst_host_rerror_rate	0	0	0.0191	0	0	0
class	Smurf.	Smurf.	Smurf.	Smurf.	Smurf.	Smurf.

The clustering results into 4 categories, according to Clustered0 src_bytes data flow source host to the destination host is 1032B,the same service connection number srv_count was 511,and the current connection the percentage dst_host_same_srv_rate connection with the same amount of the target host connection number dst_host_count is 255,the same service connection number dst_host_srv_count is 255,the same service for 100%,with the current connection the percentage dst_host_diff_srv_rate with the same target host different service is 0,the percentage of dst_host_same_src_port_rate the same source port accounted for 100%,SYN connect the percentage dst_host_serror_rate error is 0,REJ wrong connection the percentage dst_host_rerror_rate was 0,Clustered1,Clustered2,clustering results Clustered3 is expressed by the table 1 clearly shows.

Epilogue

The help software Weka3.6.13 version of the famous open source data mining, the KDDCUP99 data set "KDDCUP.data_10_percent 10 percent concentrated refuse service attack" Smurf "type, the cluster analysis was used to, will 280790 intrusion data records according to the similarity of attributes is divided into four categories, the network intrusion detection data set inherent law have a certain understanding, analyzed the intrusion data records.

References

[1] Xi Lei, Wang Feng, Wei Xiuran, Yu Hua, Zhang Hao. Design of Host-based Defense System against Smurf Attack [J]. Joumal of North China Institute of Water Conservancy and Hydroelectric Power, 2006(2):66-68

[2] Xu Yonghong, Zhang Kun, Yang Yun, Liu Fengyu. A Study on Smurf Attack and Its Countermeasures [J]. Joumal of Nanjing University of Science and Technology, 2002, 26(5):512-516.

[3] University of California.KDD Cup 1999 Data[EB/OL].http://kdd.ics.uci.edu/databases/KDDCUP99/KDDCUP99.html,1999-10-28.

Wang Xuehui,Jia Lili. Weka Makes Data Mining no Longer be Mystical [J].Computer Knowledge and Technology, 2007(5):699.

[4] Wang Xuehui, Jia Lili. Weka Makes Data Mining no Longer be Mystical [J].Computer Knowledge and Technology, 2007(5):699.

[5] Dong Xinke,Zhang Hui.Analysis and Comparison of Several Clustering Algorithms Based on Campus Card Consumption Data[J]. Application of Computer System, 2014, 23(1):158-161+183.

