Association rule mining in DoS attack detection and defense in the application of network

Jigang ZHENG[1,a*],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

Key words: Association rule; Intrusion detection; Data mining; Characteristic attribute; KDDCup99

Abstract.Association rule mining is widely used in network intrusion detection,an important feature of attribute extraction of KDDCup99 data set of denial of service attack records,for the association rules mining association rules between different attributes with the help of Weka,the intrusion detection and prevention of further study has reference significance.Mining association rules reflect the relationship between different attributes,the establishment of the intrusion detection system is effective and adaptive for the next step,has a very broad application prospects.

Introduction

With the Internet in our lives occupy an increasingly important position,very obvious,the site operators are also increasingly under serious threat in recent DoS(Denial Of Service)[1].A malicious user sends a request to the server multiple authentication to full load,and return address all requests are forged.When the server attempts to authenticate the results back to the user,it will not find these users.In this case,the server had to wait until close to the connection timeout for this connection[2].During this period the attacker will continually send bogus requests until the server overload and cannot provide normal services.

Association rule mining algorithm

Association rule mining is R.Agrawal,R.Imielinski and Swami related concepts presented in 1993,is an important field of data mining research[3][4].The so-called association rules that identify potential associations between data items described in the database to find a large number of unknown data between useful dependencies.Supermarket bar code scanner used to collect a lot of transactions,each transaction will produce a detailed list of all the information shopping transaction,association rule is derived from the analysis of the customer transaction data.Operators are always interested in what commodity to be bought together to make based on the information store layout is more reasonable,rational arrangement of commodity classification and determining the type of customer buying patterns by implementing promotional activities.

Definition1 Set $I=\{I_1,I_2,\cdots,I_m\}$ is a collection of data items, D transaction is a collection of all[5],A transaction T has a unique identifier TID.If items,transaction support items claimed T set A,also known as T transaction that contains the item set A.

Definition2 Association rules are shaped like $A\Rightarrow B$ type of implication,among them $A\subset I$, $B\subset I$,and $A\bigcap B=\Phi$.

$A \Rightarrow B$ supports the association rule is defined as: $\sup port(A \Rightarrow B) = \frac{\sup port(A \cup B)}{N} \times 100\%$,credibility is defined as: $confidence(A \Rightarrow B) = \frac{\sup port(A \cup B)}{\sup port(A)} \times 100\%$.

Support is a measure of the importance of the association rules,indicating the probability of this association rules appear in all affairs,the greater the support,the more important association rules.Credibility is a measure of the accuracy of the association rules,are drawn on the basis of association rules.

Definition 3 Support and confidence required to be greater than the threshold set by the user (ie,minimum support threshold and minimum confidence threshold),that:

$$\sup port(A \Rightarrow B) \geq \min\_\sup,$$

$$confidence(A \Rightarrow B) \geq \min\_conf$$

the strong association rules called the rules,otherwise known as the weak rule,strong association rules are useful rules researchers seek.

Research on Association Rules algorithm has appeared Apriori algorithm to generate candidate frequent itemsets,no candidate frequent itemsets FPGrowth algorithms,as well as a variety of improvements based on their algorithms.

Denial of service attacks association rule mining

Data preprocessing. Weka intelligence analysis environment full name Waikato (Waikato Environment for Knowledge Analysis),is based on Java,an open source project for data mining and knowledge discovery,and its developer is Ian H.Witten from the University of Waikato,New Zealand and Eibe Frank.After years of development,Weka is now one of the most comprehensive data mining tools,and is recognized as an open source data mining project in one of the most famous[6].In this paper,the experimental data sets from KDDCup99 dataset "KDDCUP.data_10_percent" Subset[7],this subset has 494,021 records,which recorded 391,458 denial of service attacks,accounting for 79.24%,attack classification identified as land (21 records),pod (264 records),teardrop (979 records),back (2203 records),neptune (107201 records),smurf (280790 reviews) six kinds of attack types.

KDDCup99 dataset before each record contains 41 fixed feature attributes and finally an attack type identification,select the first 13 feature attribute duration,protocol_type,service,flag,src_bytes,dst_bytes,land,wrong_fragment,urgent,hot,num_failed_logins ,logged_in,num_compromised and finally an attack attribute identifies the type of feature class,delete the remaining 28 feature attributes.KDDCup99 dataset for Excel xls file,save as CSV format,and then converted to the identification of Weka ARFF format features continuous numeric attribute grouping,were converted into discrete classification characteristic properties,shown in Figure 1.

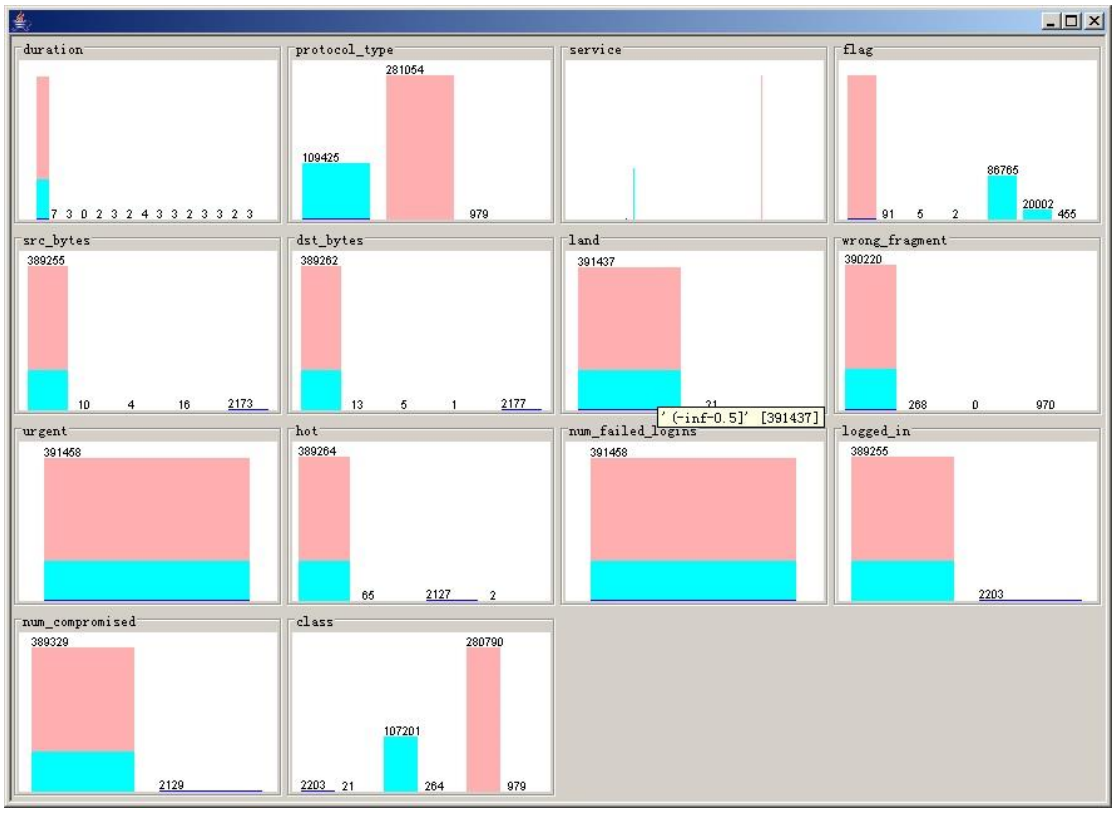

Figure. 1 Visualization map feature attributes

Association rule mining algorithm. Weka software provides Apriori,FilteredAssociator,FPGrowth,GeneralizedSequentislPatterns,PredictiveApriori,Tertius total of six kinds of association rule mining algorithm,on the basis of the experimental analysis,selection of experimental data Apriori algorithm for mining association rules.

Apriori algorithm is described as follows:[10]

In:Database D and $\min\_\sup$,

Out:Database D itemsets L,

Algorithm:

L_1 = Looking frequent two sets(D);

For k =2; $L_{k-1} \neq \Phi$; k ++

{ C_k =apriori_gen(L_{k-1});

For each transaction $t \in D$

{ $C_t = subset(C_k, t)$;

For each candidate $c \in C_t$

$c.count$++;}

L_k =($c \in C_k \mid c.count \geq \min\_\sup$)}

Return $L=\{\text{All}\ L_k\}$.

Mining results. Select the appropriate support and confidence is the key to effective mining association rules,found through experiments,along with support and increase confidence,decreases the number of rules,select the appropriate support upper and lower bounds,units of measure,decreasing iteration values have a significant impact on the results of mining,set the parameters for the "Apriori-N 20-T 0 -C 0.9-D 0.1 -U 1.0 -M 0.5 -S -1.0 -C -1",part of the mining results are as follows:

1.num_failed_logins='All' 391458==>urgent='All' 391458 conf:(1)

2.land='(-inf-0.5]' 391437==>urgent='All' 391437 conf:(1)

3.land='(-inf-0.5]' 391437==>num_failed_logins='All' 391437 conf:(1)

4.duration='(-inf-0.933333]' 391418==>urgent='All' 391418 conf:(1)

5.duration='(-inf-0.933333]' 391418==>num_failed_logins='All' 391418 conf:(1)

6.duration='(-inf-0.933333]' num_failed_logins='All' 391418==>urgent='All' 391418 conf:(1)

According to the above mining results obtained in DoS attacks association rules between features of different attributes: number of login failures num_failed_logins is 0,the number of packets urgent pressing 0,is not the same host or port land,number of login failures num_failed_logins 0,the length of time the connection duration is within 0.93 seconds.These confidence association rules were 100%,if lower confidence mining,there will be more contact occurs.

Conclusion

Characteristic properties of the network intrusion detection data set is an important indicator of the status of network intrusion,its analysis of the status quo in depth understanding of network intrusion and its laws.With the help of well-known open-source data mining software Weka3.6.2 version of KDDCup99 dataset "KDDCUP.data_10_percent" subset of denial of service attack type association rule mining,mining association rules reflect the relationship between the different characteristics of the properties,for Next to establish the effectiveness of intrusion detection system and adaptive,and has a very broad application prospects.

References

[1] ChristoPhL,Sehuba,IvanV,Krsul.Analysis of denial of service attack on TCP.Proccedings of the IEEE Computer Society Symposium on Research in Security and Privacy,1997.

[2] ZHANG Xinyou,ZENG Huashen,JIA Lei.Research of intrusion detection system dataset KDD CUP99 [J].Computer Engineering and Design,2010,31(22):4809-4816.

[3] Agrawal R,Srikant R.Fast Algorithm for Mining Association Rules.In Proceeding 1994 International conference Very Large Data Base(VLDB’94).Santiago,Chile,Sept,1994:487-499.

[4] Jiawei Han,Micheline Kamber.Data Mining Concepts and Techniques[M].Beijing:Machinery Industry Press,2007:146-168.

[5] Jiawei Han,Micheline Kamber.Data Mining Concepts and Techniques,Second Edition[M].Bei Jing:Machinery Industry Press,2007:254-255.

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

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

[8] Pang-Ning Tan,Michael Steinbach,Vipin Kumar.Data Mining Introduction.Bei Jing:The people post and Telecommunications Press,2006:206-216.