IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-10, Oct 2018 E-ISSN: 2347-2693

An Effective Feature Extraction through Fourgram Scheme for Long Payloads in Network Intrusion Detection Systems

Abdul Rustum Ali[1*], K N Brahmaji Rao[2]

[1,2]Dept. of CS & SE, A.U. College of Engineering, Andhra University, Visakhapatnam, AP, INDIA

[*]*Corresponding Author: rustumali202@gmail.com, Tel.: +91-8341424523*

Available online at: www.ijcseonline.org

Accepted: 16/Oct/2018, Published: 31/Oct/2018

***Abstract*—** Now a day's security is very global issue in Network based systems. The rate of cyber terrorism has increased day by day and it put national security under risk. In addition, network attacks have caused several damages to different sectors (i.e., individuals, economy, enterprises, organizations, and governments). Network Intrusion Detection Systems are giving the solutions against these attacks. NIDS always need to improve their performance in terms of increasing the accuracy and decreasing false alarm rates. Feature selection gives the ranking to the data set attributes and it can help for selecting the most important features from the entire set of data. In the previous researches feature selection selects the irrelevant and redundant features. These are causes of increasing the processing speed and time. An efficient feature selection method eliminates dimension of data and decrease redundancy and ambiguity. In the present network intrusion detection systems working with the long payload features are not easy tasks because many machine learning algorithms can't handle these long payload features. Some of the Network Intrusion Detection Systems are not process these long payload features. To solve this problem, a new methodology called feature extraction through Fourgram technique has been proposed. The long payload features are processed these proposed technique and prepared to be implemented in machine learning algorithms and the results were carried out on ISCX 2012 data set. The designed feature selection system has shown a very good improvement on the performance using different metrics like Accuracy, F- Measure etc.,

Keywords—NIDS,Feature Extraction,Fourgram Scheme, Long Payload Features, Dataset, Dictionary Building

I. INTRODUCTION

Generally, in real time network traffic, the payload features are long and are of different data type. It becomes extremely difficult to any intelligent systems like machine learning systems to handle such long payload features. In the present work, lot of experiments was carried out on ISCX data set. In the literature, researchers used different IDS data sets for testing their models. However, in this paper the ISCX 2012 intrusion detection data set is used for better comparisons in the results because some of the earlier and recent works [2] used this same data set. This data set has been generated by the Information Security Centre of Excellence (ISCX) at the University of New Brunswick in 2012 [2]. The data set consists of real traces analyzed to create profiles for agents that generate real traffic for FTP, HTTP, SMTP,SSH, IMAP, POP3, etc., [2]. The generated data set contains various features that includefull packet payloads in addition to other relevant features such as total number of bytes sent and /or received. This ISCX Data set consists of different types of features like numeric, alpha numeric, date, time,categorical and strings. Usually the packet header information is represented by a combination of these above types, but the payload features are usually represented by long string values which contains very long strings that makes it very difficult for any machine learning algorithms to deal with to address this problem, encoded schemes have been chosen to encode these features by using Fourgram techniques. Fig.1 illustrates the main steps of the feature extraction process that is employed to extract features using a proposed scheme. The present approach and algorithms are very similar to the procedures in but in the present study, along with Bigram and Trigram scheme Fourgram approach is also studied and experimented on the same data set. This Fourgram technique is used with payload features to investigate if the payload features contain informative features or not. It is opted to do this since most research ignores these features due to their long strings, which makes them difficult to utilize in machine learning.

The rest of the paper is organized as follows: In Section II, A view on the related work of the topic is presented,Section III describes the description of the ISCX 2012 Data Set, Section IV explains the proposed methodology Fourgram for feature extraction,and Section V explains the results of applying the Fourgram scheme on the ISCX 2012 Data set. The conclusion of the work in Section VI and the future work is presented in Section VII.

II. RELATED WORK

Lot of contributions are made available in the literature by various researchers to build more efficient NetworkIntrusion Detection Systems. In the very recent past, TarfaHamed, Rozita Dara, Stefan C. Kremer [1]designed a Network intrusion detection system based on recursive feature addition and bigram technique, in whichthey proposed a new feature selection method called RecursiveFeature Addition (RFA) and bigram technique.In fact, this work gives motivation to this present study.In this section, some of the papers that were cited in [1] were also studied. Apart from the above work, Studies have been conducted on applying feature selection to improve the IDSperformance. In [1], the authors applied the intra class correlation coefficient and interclass correlation coefficient to attain a class specific subset of features.The interclass and intra-class correlation coefficients wereused to measure the validity and the reliability of features respectively. The authorsShiravi A, Shiravi H, Tavallaee M, Ghorbani AA [2] tested their model on the ISCX 2012 data set. They observed that the above combination between interclass and interclass correlation coefficients led to an increase in the detection rate and to a decrease in both execution time and false alarm rate. However, their work did not deal with the scarcity of data and interdependent features as the authors in did in their work.

In other studies [2], the authors opted to build their intrusion detection system based on the normal traffic todetect unseen intrusions using the ISCX 2012 data set. The authors employed a one-class Support Vector Machine (SVM) classifier to learn http regular traffic attributes for an anomaly detection task. Their approach involved extracting appropriate attributes from normal and abnormal traffic. The system generates an alert if it finds any deviation from the normal traffic model. The authors stated that they obtained 80% accuracy and 8.6% false alarm rate in detecting attacks on port 80. Authors in stated that their work differs from the work in [3] in dealing with normal and attack data instead of dealing with normal data only. As the present work imitates the work in with respect to experiments on the data set, both normal and attack data is considered in this work.

Zero-day attacks have made known to be intricate to alleviate their damage due to the lack of information [4].For this reason, there is always a need to protect against these zero-day attacks before they causeenormousdamage to networks. These attacks are also called "zero-day misuses" [5][6]. As just mentioned,Zero-day assaults have appeared to be hard to lighten their harm because of the absence of data [7][8].Consequently, there is dependably a need to protect against these zero-day assaults before they maketremendous harm to the systems. Information mining is a method that can be utilized with interruption identification to distinguish trademark designs from the information included in that portray framework and clientconduct [9][10], and preferably, cases of pernicious action. Machine learning calculations have been utilized broadlywith interruption discovery to improve the precision of identification and making a safe model for the IDSagainst zero-day assaults or novel assaults [11][12].

To construct quick and exact IDS, it is essential to choose to enlighten highlights from the information. Highlightdetermination has demonstrated its capacity to diminish calculation requests, over fitting, display size andincrement of the exactness [10][13]. The trouble that faces an engineer assembling these sorts of frameworks is theshortage of assault illustrations which can be utilized to prepare a learning machine to manufacture a model foridentifying that specific assault. Indeed, even powerful machine learning calculations battle when there are acouple of illustrations, or unequal cases and substantial quantities of highlights. The accessible useful highlightslikewise influence the execution (that is the more the better). Past IDS regularly ignored the payload highlightsin spite of the fact that they contain some helpful data [14][15]. In this way, we chose to use the payloadhighlights and concentrate helpful data for ID purposes. Keeping in mind the end goal to enhance theidentification capacity of the framework, The Bigram and Trigram strategy was utilized to encode the payloadhighlights into a shape that can be utilized as a part of machine learning calculations. The Bigram system is a set up procedure particularly in Deep Packet Inspection (DPI) and has been contemplated for quite a long time [16][17].Be that as it may, in this system, another mix of utilizing highlight choice, the Bigram procedure and theapplication to this specific issue (interruption recognition) is exhibited. Experiments were carried out to address theissue of interruption discovery harder by concentrating on "zero-day assault" situation. With a specific end goal toreproduce this, it is deliberately fabricated a learning machine utilizing little quantities of cases and extensivequantities of highlights. The reason for that is to check on the off chance that can be even now be distinguishedassaults with an informational index with the above attributes.

As outlined in this section on some of the studies in the area of NIDS, it may be concluded that in spite of decadesof research in this area, handling long payload features on the network traffic still remains as challenge.

III. About ISCX 2012 data set

In this section,we will explain the ISCX 2012 data set. This dataset has been generated by the Information Security Centre ofExcellence (ISCX) at the University of New Brunswick in 2012(Shiravi et al., 2012). The data set involves real traces analyzedto create profiles for agents that generate real traffic forHTTP, SMTP, SSH, IMAP, POP3, and FTP. The generated datasetcontains different features

including full packet payloadsin addition to other relevant features such as total numberof bytes sent or received. The full data set has been capturedin a period of seven days (From Friday June 11th at 00:01:06sto Friday June 18th at 00:01:06s) and involved more than onemillion network trace packets and 20 features, and every dataexample has been labelled as one of two classes (normal orattack) (Yassin et al., 2014).The ISCX data set has acquired thesecurity community's attention and become a benchmark dataset for intrusion detection research purposes due to its realistictraffic, labelled connections, and its multiple attackscenarios (Shiravi et al., 2012). The data set has been designedto overcome the technical limitations of other intrusiondetection datasets, and to prepare network traces by capturingcontemporary legitimate and intrusive network behaviorsand patterns (Tan et al., 2015).

Table 1: Description of ISCX 2012 Dataset

Day	#features	#examples	#normal	#attacks
June 11th	19	378,667	378,667	0
June 12th	19	133,197	131,110	2087
June 13th	19	110,588	632	109,956
June 14th	18	171,388	167,594	3786
June 15th	20	572,410	534,830	3840
June 16th	20	523,160	523,134	26
June 17th	19	398,516	393,181	5335

IV. PROPOSED FOURGRAM SCHEME FOR EFFICIENT FEATURE EXTRACTION

The Bigram technique is an established technique especially in Deep Packet Inspection (DPI) and has beenstudied for decades [18]. However, in this paper, not only a new combination of using the Bigram and Trigram techniques arefeature selection is experimented as in [16], but also a new proposed Fourgram technique and the application ofthese scheme over the long payload features ispresented.

In the proposed methodology the initial step in the feature extraction process for all payload features is construction of the dictionary. Consecutively, to extract the feature vector for each payload feature, a dictionary need to be built that contains all the Fourgramsrespectively. Fourgram scheme have beenproposedso create the dictionaryto perform various experiments. This concept is explained in this paper by taking various examples as it is done in [16]. The featureextraction process in this approach is depicted in figure 1, which is similar to the approach in [16].

The long payload features are encoded using Fourgram technique. Fig. 2 illustrates the main steps of thefeature extraction process that is employed to extract features using Fourgram techniques. The Trigramscheme has already been implemented in [16], in the present work the Trigram approach is once again experimented on slightly different data and along with it the Fourgram approach has also been experimented and results are noted and are presented for comparative analysis. ThisFourgram technique is used with payload features to investigate if the payload features include informative features or not.

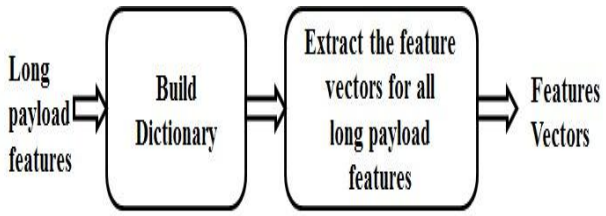


Fig. 1: Feature extraction process for ISCX dataset using Fourgram technique

The detailed steps for the dictionary construction areshown in the following Algorithm 1. The only difference betweenthe Trigramand Fourgram are in Trigram scheme three adjacent words are takenfrom a feature whereas in Fourgram four adjacent words are taken. Hence a common algorithm is presented below. Themethodology of constructing Trigrams dictionary andFourgrams dictionary are similar. The output of thealgorithm will be the dictionary with Fourgrams respectively.

Algorithm 1: Dictionary Construction for Long Payload Features

Step-1: Input the long payload features
Step-2: Initialize the dictionary "D", with empty
Step-3: Take one feature from the long payload
Step-4: Take one Fourgram from the feature
Step-5: Check if the dictionary consist this Fourgram already
Step 6: If dictionary D does not contain Fourgram then add feature to dictionary D
Step-7: Repeat step 4 to 6 till there is no possibility of new Fourgram from the features
Step-8: Repeat the procedure till all the payload features are encoded as Fourgrams and added to dictionary D
Step-9: Output: Dictionary D

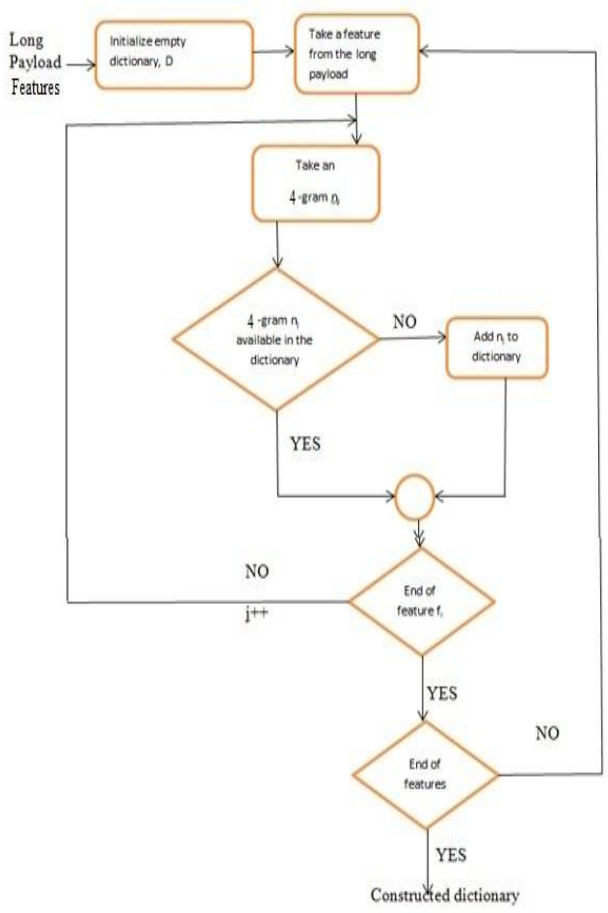


Fig.2 Dictionary construction stage during feature extraction for ISCX Dataset using Fourgram Technique

Figure 2 shows the block diagram of dictionary construction. Now, this dictionary which has been built with Fourgrams can be used by any feature extraction algorithm more effectively even from the long payload features. The generated Fourgrams can also be handled by any machine learning algorithms. In the present work, feature selection process was done by using one of the most widely cited machine learning algorithm, Support Vector Machine (SVM) to study the results out of the proposed methodology. Weka GUI v3.8 has been used for experiments. The next step is the feature vector extraction for the long payload features which is outlined in the following figure 3. The approach in [2] has been followed in the present work for the task of feature vector extraction. The feature vector extraction step is also explained in more detail in the below Algorithm 2.

Algorithm 2: Feature Vector Extraction for Long Payload Features with Fourgrams

Step-1: Input the long payload features and constructed Fourgram Dictionary

Step-2: Initialize all feature vectors to zero

Step-3: Take one string at a time

Step-4: Take one Fourgram

Step-5: Find the index of Fourgram from the dictionary

Step-6: Increment the location counter in the feature vector

Step-7: Repeat till there is no possibility of findingFourgram from the inputpayload feature

Step-8: Finish feature vector i, and proceed to feature vector i+1

Step-9: Stop the process if all there no feature vector is left to be processed

Step-10: Output the feature vectors

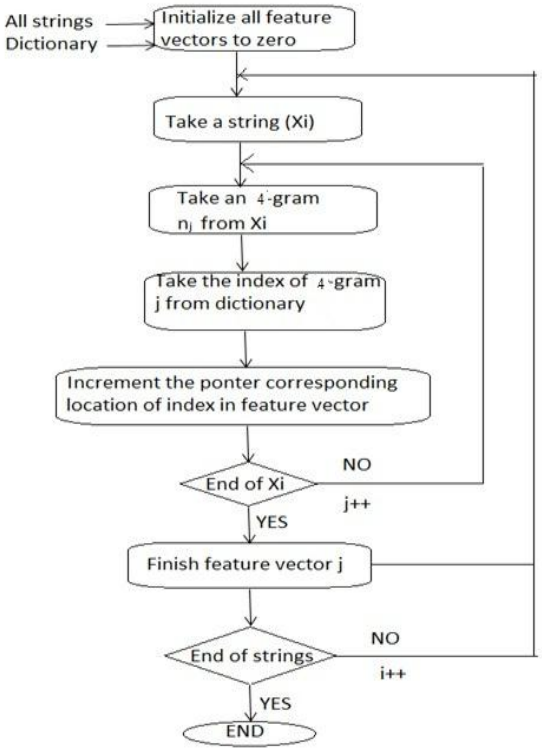


Fig.3 Feature vector extraction for ISCX data set using Fourgram technique

V. EXPERMENTS AND RESULTS

The experiments were carried out on the available ISCX2012 data set. The result of Algorithm 2 is the featurevectors for all the payload string features from this ISCX2012 data set. After this step, the number of features canexpand from hundreds to thousands and even to millions.This actually motivated to propose a new scheme ofencoding, called "Fourgram", which somehow reduces thiscount. To explain the proposed scheme more clearly, ademonstration is made with small example on how thepayload features are converted to Fourgramsaccording tothe proposed methodology. To avoid the complexity,onlythree training instances have been taken as in [1] butdifferent payload features are taken. The content of theeach payload feature is different.

Generation of Bigrams:
Suppose the actual three payload features are: "ko9vr","sfruw" and "fvobga" respectively then by feedingthose features to the dictionary generation according toAlgorithm 2, the resulting dictionary will consist ofthirteen (13)bigram words as follows: ko | o9 | 9v | vr | sf |fr | ru | uw | fv | vo | ob | bg | ga. The redundant Bigrams areexcluded from the list.

By applying the feature vector extraction on the giventhree payload features according to Algorithm 2, theresulting bigram representation for each of the abovethree features will be as presented in Table 2.

Table 2: Bigram representation for the three payload features in the given example

Original payload	ko	o9	9v	vr	sf	fr	ru	uw	fv	vo	ob	bg	ga
ko9vr	1	1	1	1	0	0	0	0	0	0	0	0	0
sfruw	0	0	0	0	1	1	1	1	0	0	0	0	0
fvobga	0	0	0	0	0	0	0	0	1	1	1	1	1

Generation of Trigrams:
Here, the three payload features that are taken as examples are same as taken for Bigram generation. The resulting Trigram dictionary consists of ten (10) Trigram as follows:
ko9 | o9v | 9vr | sfr | fru | ruw | fvo | vob | obg |bga. Theredundant Trigrams are excluded from the list.

By applying the feature vector extraction on the giventhree payload features according to Algorithm 2 and theresulting Trigram representation for each of the abovethree features will be as depicted in Table 3.

Table 3: Trigram representation for the three payload features in the given example

Original payload	ko9	o9v	9vr	sfr	fru	ruw	fvo	vob	obg	bga
ko9vr	1	1	1	0	0	0	0	0	0	0
sfruw	0	0	0	1	1	1	0	0	0	0
fvobga	0	0	0	0	0	0	1	1	1	1

Generation of Fourgrams:
Here, the four payload features that are taken as example are same as taken for Bigram generation.
The resulting Fourgram dictionary consists of seven (7) four gram as follows:
ko9v | o9vr |sfru | fruw | fvob |vobg |obga. The redundant Fourgrams are excluded from the list.

By applying the feature vector extraction on the given three payload features according to Algorithm 2, the resulting four gram representation for each of the above three features will be as depicted in Table 4

Table 4: Fourgram representation for the three payload features in the given example

Original payload	ko9v	o9vr	sfru	fruw	fvob	vobg	obga
ko9vr	1	1	0	0	0	0	0
sfruw	0	0	1	1	0	0	0
fvobga	0	0	0	0	1	1	1

The three payload features in this example were converted to Bigram, Trigram and Fourgram features as in Table 2, table 3 and table 4 respectively. The table's header represents the standard feature vector for all the payload features in the taken example. The payload features appear in the table as per the order of their presentation to the feature extraction algorithm, that is the first row corresponds to the firstpayload feature discovered during dictionary generation and second row corresponds to the second payload feature discovered during dictionary generation and so on.

After all this pre processing, in order to prepare theresulting data set for feature selection, a pre ranking stephas been conducted for the features using a selection algorithm that is the well known gain ratio feature selection. The reason for thisranking is to make sure that the reducedfeatures are stillinformative andhas far; still contained distractingfeatures as well. Hence a very fast feature selectionmethod is used to pre rank the features, which is calledgain ratio feature selection. By using thismethod, it is easy to rank the features in order to performcontrolled experiments by manipulating relevant and irrelevant features.

Because the current number of features is very huge and the feature selection is really a time consuming activity, hence a small subset of 300 features from the originalfeatures are taken for experimental purpose. From the obtained ranked features list from the gain ratio method, a smaller subset of this ranked list is extracted in such away that it consists of one portion of the top ranked features and nine portions of lowest ranked features. The considered portion size in this work is 30 features; therefore the subset size is obviously 300 features in total.This way of selection is intentional because to test theefficacy of the proposed system, experiments wereconducted on the data set, in which 90% of the totalfeatures are bad features and only 10% are good features.The small subset of features is taken as sample from thehuge set of total features to save time and computationaleffort.

Those resulting 300 features are used to generate datasetsof size 30, 60, 150, and 500 examples respectively. In order to simulate "zero-day" attacks, the datasets havebeen chosen to be small in terms of number of examples.Those data sets were generated as balanced data sets (i.e.,equal number of normal and attack examples). Differentnumbers of examples were used to monitor the behaviorof feature selection with each size of dataset.

To observe the effect of including the payload features toimprove the detection accuracy, a very importantexperiment has been conducted. The well known machinelearning algorithm, the SVM's classification was used tofind out the accuracy and F-measure on the ISCX 2012data sets in two cases, first without (Bigram /Trigram / Fourgram features) the payload features and second with (Bigram / Trigram / Fourgram features) the payload features. Theperformance metrics have been measured before and afterconverting the payload features into Bigram, Trigram and Fourgram features and applying feature selection.

The main objective of this experiment is to show thatthese payload features include important and usefulinformation in improving the detection accuracy. Becauseof the inability of handling the long payload futures, manyof the previous researchers excluded payload featuresfrom the original set of features before looking forintrusions. The feature selection method was applied onthe four generated data sets of ISCX 2012 and presentedthe maximum obtained accuracy and F-measure as shownin Table 5, Table 6and Table 7 respectively.

Table 5: Performance metrics without and with Bigram features on the ISCX data sets

Data Set	Without Bigram features		With Bigram features	
	ACC	F-measure	ACC	F-measure
30 examples	67.20%	67.20%	78.60%	78.00%
60 examples	79.00%	79.00%	89.60%	89.60%
150 examples	79.50%	79.50%	89.80%	89.70%
500 examples	83.89%	83.89%	93.90%	93.90%

Table 6: Performance metrics without and with Trigram features on the ISCX data sets

Data Set	Without Trigram features		With Trigram features	
	ACC	F-measure	ACC	F-measure
30 examples	67.20%	67.20%	77.60%	77.00%
60 examples	79.00%	79.00%	88.70%	89.21%
150 examples	79.50%	79.50%	89.20%	89.20%
500 examples	83.89%	83.89%	93.01%	92.90%

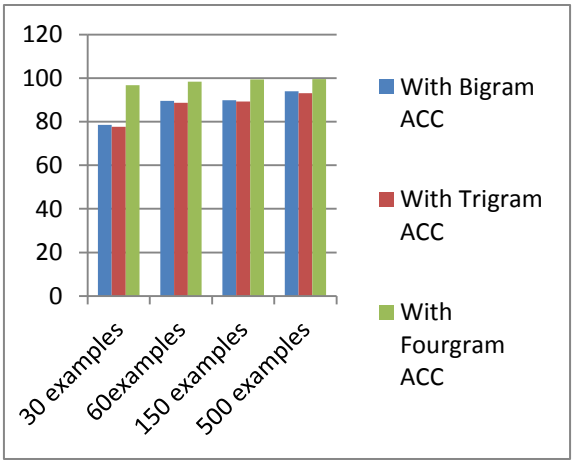
Table 7: Performance metrics without and with Fourgram features on the ISCX data sets

Data Set	Without Fourgram features		With Fourgram features	
	ACC	F-measure	ACC	F-measure
30 examples	67.20%	67.20%	99.66%	96.77%
60 examples	79.00%	79.00%	98.33%	98.36%
150 examples	79.50%	79.50%	99.33%	99.32%
500 examples	83.89%	83.89%	99.56%	99.56%

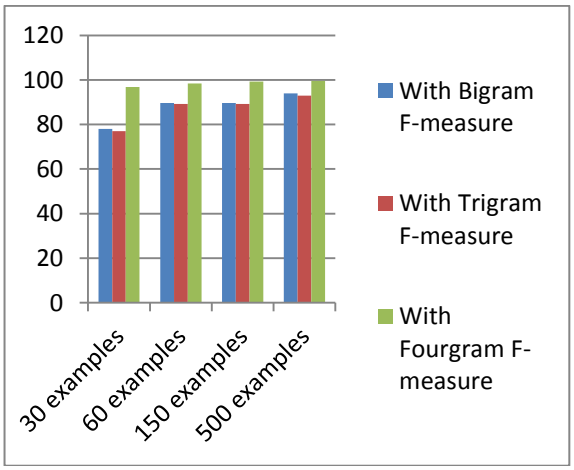
It may be clearly observed from the above table that thereis a clear improvement in the accuracy and F- Measurewhile considering the Bigram, Trigram and Fourgram features. Theclassifier's performance on each data set with the payloadfeatures removed and included from that data set has beenstudied. Columns 4 and 5in the tables represent themaximum obtained performance from the classifier afterincluding the payload features, expanding them using theFourgram technique and applying feature selectionalgorithm on the resulting data set. It may be noted that here is a considerable improvement around 10% afterapplying classifier and feature selection algorithm on the Fourgram features compared to the performance ofthe SVMs classifier on the same data sets without Fourgram features.

One of the major contributions in this work iscomparative analysis between the Bigram, Trigram and Fourgram onthe same data set of long payload features. It may also beconcluded after the studies on the experimental results that Bigram, Trigram and Fourgram features inclusion haveimproved the performance of the classifier in terms ofaccuracy and False Measure. Even though there is aminute improvement using Bigram features over Trigramfeatures in some instances, there will be a considerablereduction in the

computational overhead. The below twographs, Graph1 and Graph2 will give a clear picture on thealmost similar results with Bigram, Trigram and Fourgram withrespect to Accuracy and F-measure as performancemetrics.



Graph1: Accuracy achieved over Bigrams, Trigram &Fourgrams on the ISCX datasets



Graph2: F-measure achieved over Bigrams, Trigrams & Fourgrams on the ISCX datasets

Hence after various observations from the experimentalresults it is suggested to use fourgram features to includelarge payload features instead of trigram in any intelligentfeature selection system to reduce the computationaloverhead.

VI. CONCLUSION

In this paper, to handle long payload features, three encoding schemes, where two of them are available in the literature and other is a new one, have been studied and experiments were carried on the proposed method on the ISCX 2012 data set. The data set has been prepared for intrusion detection by processing the long payloads using Fourgram technique. The most of the previous techniques are performing feature selection have focused only on the statistical properties of packets, an attempt was made to also try and extract useful features from the long payloads using a Fourgram technique. Apply this method produced a high dimensional data set with thousands of features from the given payload features. It is observed that the data set with large numbers of features and relatively few numbers of examples is always useful in testing the resilience of the system against over fitting. The experimental results are encouraging in drawing useful conclusions such as to recommend fourgram features to include large payload features instead of trigram in any intelligent feature selection system to reduce the computational burden.

VII. FUTURE EXTENSION

By proposing the new effective feature selection method, classification algorithm and the construction of efficient data set will give us the accurate and best results while processing the data through NIDS systems.

ACKNOWLEDGEMENTS

Authors want to express their special thanks to Dr.ArashHabibiLashkari, Research Associate R&D Manager, Canadian Institute for Cyber security (CIC) and other team members at CIC for the generosity in responding to the request of the authors for various data sets including ISCX 2012. The data sets provided by CIC are very useful in this current work in carrying out various experiments and draw useful conclusions.

REFERENCES

[1] TarfaHamed, Rozita Dara, Stefan C. Kremer, Network intrusion detection system based on recursive feature addition and bigram technique, International journal of computers & security, Vol. 73, pp.137–155, 2018.

[2] Shiravi A, Shiravi H, Tavallaee M, Ghorbani AA, "Toward developing a systematic approach to generate benchmark datasets for intrusion detection", International journal of Computer Security, Vol.3,No.31,pp.357-374,2012.

[3] Garcia LP, de Carvalho AC, Lorena AC, "Effect of label noise in the complexity of classification problems",International journal of Neurocomputing", Vol.19,No.16,pp.108-119,2015.

[4] Beigi EB, Jazi HH, Stakhanova N, Ghorbani AA,"Towards effective feature selection in machine learningbased bonnet detection approaches", In the proceedings of the 2014 IEEE conference on Communications and Network Security,pp.247–255, 2014.

[5] Bolon-Canedo V, Snchez-Maroo N, Alonso-Betanzos A, Bentez J, Herrera F, "A review of microarray datasets and applied feature selection methods", International journal of Information Sciences",Vol.5,No.42,pp.111-135,2014.

[6] Beniwal S, Arora J, "Classification and feature selection techniques in data mining", International journal of engineering and Research and technology, Vol.1, No.6, pp.1-6, 2012.

[7] Fahad A, Tari Z, Khalil I, Habib I, Alnuweiri H, "Toward an efficient and scalable feature selection approach for internet traffic

classification", International journal of Computer Networks",Vol.9.No.57,pp.2040-2057,2013.
[8] Aghdam MH, Kabiri P, "Feature selection for intrusion detection system using ant colony optimization", International journal of network security, Vol.3, No.18, pp.420-432, 2016.
[9] Bolón-Canedo V, Sánchez-Maroño N, Alonso-Betanzos, "A review of feature selection methods on synthetic data", International journal of information systems, Vol.3, No. 34, pp.483-519, 2013.
[10] Sahu SK, Sarangi S, Jena SK, "A detail analysis on intrusion detection datasets"In the proceedings of the 2014 IEEE International conference on Advance Computing Conference (IACC), pp.1348-1353, 2014.
[11] Mancini LV, Di Pietro R, "Intrusion Detection Systems", International journal of Springer, Vol.5, No.9,pp.513-524, 2008.
[12] Ambusaidi MA, He X, Nanda P, Tan Z, "Building an intrusion detection system using a filter-based feature selection algorithm", International journal of Computer science and applications, Vol.10, No.65, pp.2986-2998, 2016.
[13] Abou El Kalam A., Gad El Rab M., and Deswarte Y, "A model-driven approach for experimental evaluation of intrusion detection systems, International journal of Security Communication Networks, Vol.7, No.14, pp.1955–1973, 2014.
[14] Mell P, Hu V, Lipmann R, Haines J, Zissman M, "An overview of issues in testing intrusion detection systems", Technical Report, NIST IR 7007, National Institute of Standard and Technology, USA, 2003.
[15] NiccolòCascarano, Luigi Ciminiera, FulvioRisso, "Improving cost and accuracy of DPI traffic classifiers", In Proceedings of the ACM Symposium on Applied Computing (SAC '10) ACM, New York, NY, USA, pp.641-646, 2010.
[16] Laurent Bernaille , Renata Teixeira , Ismael Akodkenou,Augustin Soule , KaveSalamatian, "Traffic classification on the fly", Journal of ACM SIGCOMM Computer Communications, Vol.36, No.2, pp.145-158, 2006.
[17] Zhang M, Wang L, Jajodia S, Singhal A, Albanese M, "Networkdiversity: a security metric for evaluating the resilience ofnetworks against zero-day attacks", International journal of transformation of information science Vol.5, No.11, pp.1071–1086, 2016.
[18] Chang C-C, Lin C-J. LIB, "SVM: a library for support vector machines", International journal of Intelligent System Technologies, Vol.3, No.2, pp.215-226, 2011.

Authors Profile

Mr. Abdul Rustum Ali, obtained his B.Tech in Computer Science and Engineering from Krishna University. He is currently pursuing M.Tech in Computer Science and Technology in AUCE(A), Andhra University Visakhapatnam. His main area of intrest in Machine learning.

Mr. K.N.Brahmaji Rao, obtained his M.Sc. in Mathematics from Andhra University, M.Phil in Mathematics from Madhurai Kamaraj University, M.Tech in Computer Science and Technology with specialization in Artificial intillegence and Robotics, Andhra University. He is Pursuing Ph.D. in Computer Science and Engineering, Andhra University. He is currently working as guest faculty in the department of Computer Science & Systems Engineering, AUCE(A), Andhra University since 2016. His main research work focuses on Text Based Mining and Machine Learning. He has 18 years of teaching experience.

