Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 45 (2015) 506 – 513

Procedia Computer Science

International Conference on Advanced Computing Technologies and Applications (ICACTA-2015)

A Novel Approach to Deep Packet Inspection for Intrusion Detection

Thaksen J. Parvat[a], Pravin Chandra[b]

G.G.S. Indraprastha University,
USET[a], USICT[b], New Delhi, and Dwarka-78. INDIA
pthaksen.sit@sinhgad.edu, chandra.pravin@gmail.com

Abstract

Intrusion detection and prevention is an integral part of Deep packet inspection. There are many methods and algorithms to detect signatures; each has its own merits and limitations. The measure of this time, accuracy and space requirement. All methods and algorithms are enhanced with technology revolutions. The development in anomaly and misuse detection in this decade is crucial as web services grow vast. Managing secure network is a challenge today. The objectives vary according to the infrastructure management and security policy. There are various ways to identify payload traffic using DPI, Network security, privacy and QoS. The functions of DPI are protocol detection, anti-virus, anti-malware and IDS/IPS. The detection engine may support by a signatures or heuristics. Most of the algorithms do training and testing separately, it takes approximately double time. The paper suggests a new model to improve performance of deep packet inspection for Intrusion detection system.

Peer-review under responsibility of scientific committee of International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).

Keywords: Intrusion Detection, Packet Inspection, Security, Performance..

Corresponding author. Tel.: +91-9881920029; fax: +91-2114 -278 304.
E-mail address: pthaksen.sit@sinhgad.edu

Peer-review under responsibility of scientific committee of International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).
doi:10.1016/j.procs.2015.03.091

1. Introduction

Network traffic inspected at layer two through seven as it moves to network host. Deep Packet Inspection (DPI) is a leading technology tool for primary services as Protocol analysis, antivirus malware protection, web filtering, Intrusion Detection and Prevention System (IDPS). It is useful for security, compliance, application recognition and billing also.[25][6] IDS is shallow packet inspection (state full packet inspection)[2] is base on signature database. IPS is an inline and automatic action when detected. The packet is five tuple (Source IP, Destination IP, Source port, destination port and network layer protocol. It is also privacy violation [26]. Any application needs efficient algorithm and high-speed pattern matching. There are hardware and software solutions to achieve better performance. The misclassification leads to increase false positive and false negative results. The correct method is important to get an accurate result. All methods are base on ports that are varying in many applications. Perfect string match is base on sequence of characters or numbers in the payload. On the basis of numerical methods, arithmetic properties of several packets are used to identify change in data between hosts. Analysis of behavior and heuristics are often combined to provide assessment capabilities [27]. The serious challenges in DPI are: how to prepare signature library, How to get a fast response with low latency, How to characterize detailed network activities.

The DPI classifier need to work at very high speed and improved classification accuracy [28]. The three packet classification tasks are single field classification, multi-field classification and deep packet classification (DPC). It does fast packet classification [5]. String search algorithms are Boyer-Moore and its modifications. The average time complexity is ranging from *O (n)* to *O (min)* [5].

Snort is publically available open-source NIDS and widely used for a host based NIDS. It can also integrate with other solutions. Snort supports many platforms but use with high-end servers in the enterprise. Snort can be build up as sniffer, packet logger, NIDS/NIPS [6]. DPI is use by some Internet service operators for measurement, management of traffic to their customers on shared line. The Internet neutrality and privacy are other important issues for debate [8].

Different methods used for DPI, NFA/DFA based methods are faster. Aprori methods are accurate but have space complexity. Signature identification methods are accurate. Port based methods are not accurate because 60% peer-to-peer traffic, and their ports are dynamic which not standard IANA port numbers [7] are.

Recently multilayered and multilevel approach is used in IDS/IPS for accuracy in detection. Same time many hardware based enhancements are there which can work up to 40 gigabyte packet processing. It includes Field Programming gate Array (FPGA), Network Processor (NP), Ternary Content Addressable Memory (TCAM)[10].

Table 1. Model Comparisons

A Generative Model	A Discriminative Model
• It learns joint probability distribution $p(x, y)$.	• It learnt Conditional probability distribution $p(y\|x)$.
• Ex. Hidden Markov model and Naïve Bayes	• Ex. Conditional random fields, Maxient classifier and Maximum entropy model
• Naïve Bayes Model is a method to classifying Single class variables independence of some Feature values.	• Maximum Entropy Model is an approach to classifying single class variables, independence of some feature values. The difference is the significance of conditional probability instead of the joint probability.
• The Hidden Markov Model is an extension to the Naïve Bayes Model for in sequence structured data [25].	• Conditional Random Fields can be understood as a sequential extension to the Maximum Entropy Model [1]

Combining expectation Algorithm and density function with k-means and k-nearest algorithm give runtime at lightweight traffic [11]. CUTE is another traffic classification technique. It is faster than any other existing methods. It replaces complex regular expressions with a simple set of rules. DPI is use for application-aware services and bandwidth efficiency. It is the base on URL analysis [13]. Further going detail can provide web usage and link level measurements [14]. Multiple patterns matching with hash is use for internet application identification and security purpose [15]. DPI is useful for detecting botnets even when it is encrypted [18]. Graphics Processor with attack signature matching gives high performance in Intrusion Detection System.

2. System development model

It will identify and divide entire network traffic into two tuple normal and abnormal. Then it will check with IDS signature database. It will divide entire dataset into two categories. While checking every record, the key attributes are verified for all possible abnormalities. If found, it will be checked for different cases for perfect detection. Here training and testing is integrated in one module that saves time. It does not skip any record as all are given as input to the system.
Probability of Normal records: x 50%, Probability of Abnormal records: y 50%.

All these are joint probability in network analysis

$P(t) = \{P(x)/P(y)) + P(y)/P(x)\}$

3. The DPI methods for IDS

All the traditional methods are base on attributes of the packet. It will select on the basis of purpose or application. Some of these are:

- Signature pattern strings.
- Standard Port numbers
- Time/flow behaviour
- Source/ destination host based
- Content statistics/behaviour

Each has its advantages and limitations [11]. We use content statistics behaviour for our work. It is based on different attributes states of normal and anomalous behaviour.

4. System Architecture

The motivation to propose this new approach is, nearly fifty percent intrusion and security violation are within the organization. The security reports generated on analyzer shows apparently. If traffic is divided into different category wise, it will be more sophisticated for further investigation also. The building block will be same for both organizational and public network traffic. It is hybrid anomaly and intrusion detection system.

We are checking packet data information first time for normal or abnormal. Second it will be checked for its attributes for which type of attacks and block if any. In modular, layered and sequential searching the time complexity and space complexity is exponential $O(n^m)$. It is check for all class of attribute set. It proposes to check attributes once and identify for whether DoS, Probe, R2U, U2R or Unknown. So it's time complexity is in the range of $O(n/2)$. its worst case complexity is multiple of number pattern and text count O(n). It is acceptable till n^2. The basic searching algorithms complexity is given in Table No 3.

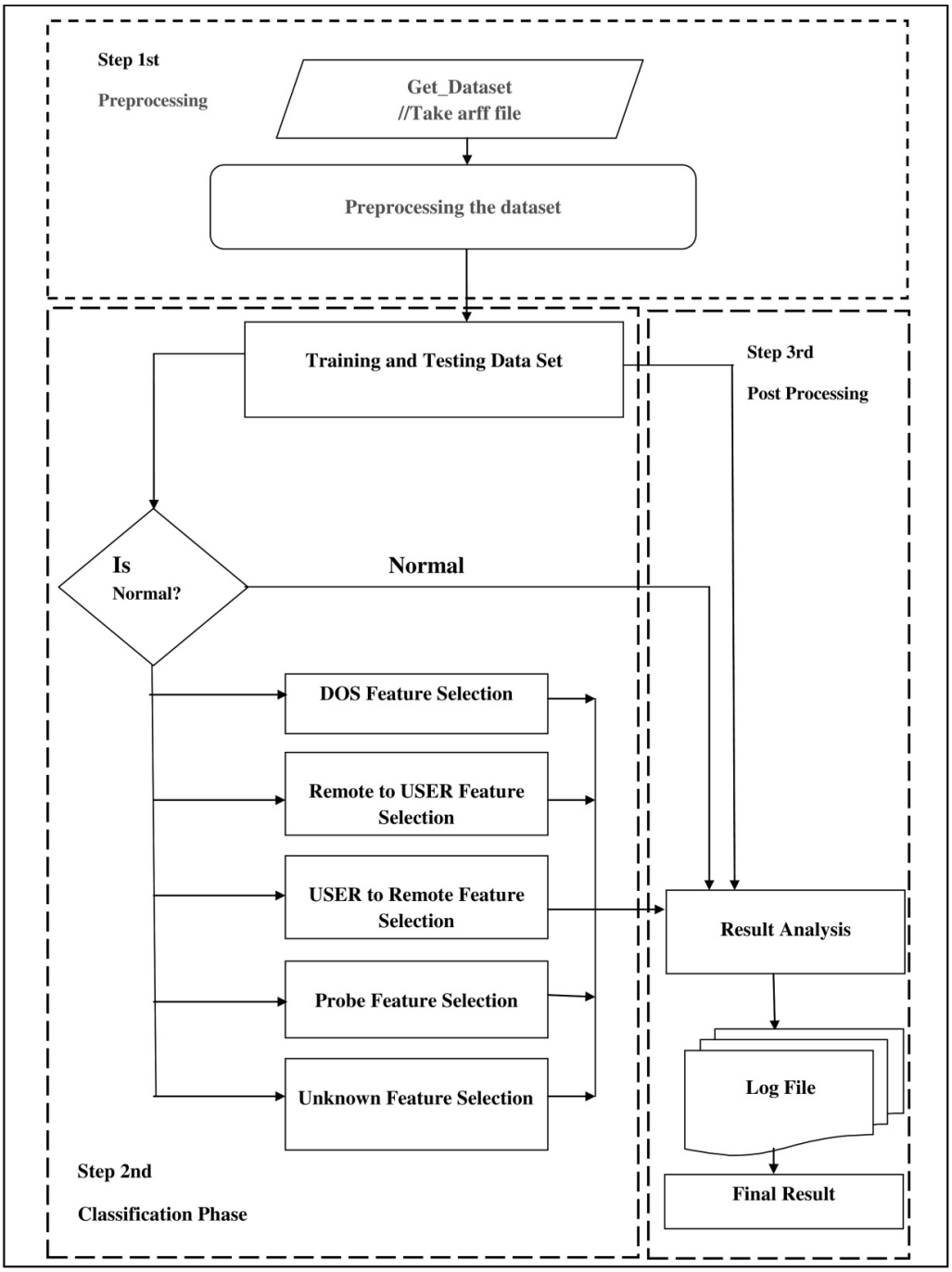

Fig. 1. Architecture and System Flow

Table No. 2 Searching Algorithms Complexity

Algorithm	Best case	Worst case	Average Case
Brute force	O(mn)	$O(n^2)$	O(mn)
Sequential Search	O(1)	O(n)	O(n)
Optimal Missmatch	O(mn)	O(m+n)	O(mn)
Boyer Moore	O(n/m)	O(nm)	O(n)
Our Algorithm	O(n)	O(n/2)	O(n/2)

Algorithm 1: Training

```
// Select Data Set
For(int i=mStart;I<mLength;i++)
{
If((Case 1) || (Case 2) || (Case 3) || (Case 4) ||……..|| (Case 11)))
{
/* By using this conditions or Cases we separate out Anomalies and Normal data sets
There are following Cases is present in this algorithm */
Count++;       //Counter for the abnormal records
If(class=ipsweep,portsweep ,nmap, satan ,mscan, saint)
{
Probes++          //Counter for the probes records
}
If(class=black,land,Neptune,pod,smuf,teardrop,apache2,mailbomb,processtable,udpstorm)
{
Dos++                //Counter for the DoS records
}
If(class=buffer_overflow,loadmodule,perl,rootkit,httptunnel,ps,worm,xterm)
{
U2R++                //Counter for the U2R records
}
If(class=ftp_write,guess_password,imap,multihop,phf,spy,warzclient,waremaster,named,sendmail,snmpgetattack,
snmpgus,sqlattack,xclock,xsnoop)
{
R2L                        //Counter for the R2L records
      } }
```

Algorithm 2: Testing

\\ Cases for the 1st If

Case 1

Attribute	4	5	12	23	26	29	30	33	34	38	39	40
Value	S0/REJ	0	0	>=1	<=1	<=1	>0	<100	<=1	<=1	<=1	<=1

Case:2

Attribute	2	3	4	5
Value	Icmp	eco_i/ecr_i	SF	<30/>500

Case 3

Attribute	4	5
Value	RSTR/SF	54540

Case 4

Attribute	4	14
Value	RSTO	0

Case 5

Attribute	4	5
Value	OTH	http

Case 6

Attribute	4
Value	RSTOS0

Case 7

Attribute	4
Value	SH

Case 8

Attribute	3	4	5
Value	Private	SF	!=105

Case 9

Attribute	3	4	5
Value	Other	SF	<10/==334

Case 10

Attribute	2	3	4	5	6
Value	tcp	ftp	SF	>1000	<2455

Case 11

Attribute	3	4	5
Value	ftp_data	SF	334

5. Experiments and Results

We use KDD NSL whole Dataset and 20% Dataset for experiment. It has 41 attributes and 38 attacks. We have implemented our algorithm with JDK 1.7 and NetBeans 8.0 IDE. We compare our results with WEKA explorer results of ByesNet, NaiveByes, Decision Table and Trees J48 to our project. These classifiers use training and testing phase on Dataset. WEKA use *.arff* files, Other format files like *.csv* need to be converted in *.txt.*

Following Table 3. Summarize results of various classifiers in confusion matrix format. The 20 % dataset includes total 25192 records.

Table 3. Confusion Matrix

Category	True Normal	True Anomaly	False Normal	False Anomaly
BayesNet	13330	10996	119	747
NaiveByes	12272	10298	1177	1445
Decision Table	13372	11563	77	180
Tree J48	13389	11692	60	51
Our Algorithm	13372	11443	77	300

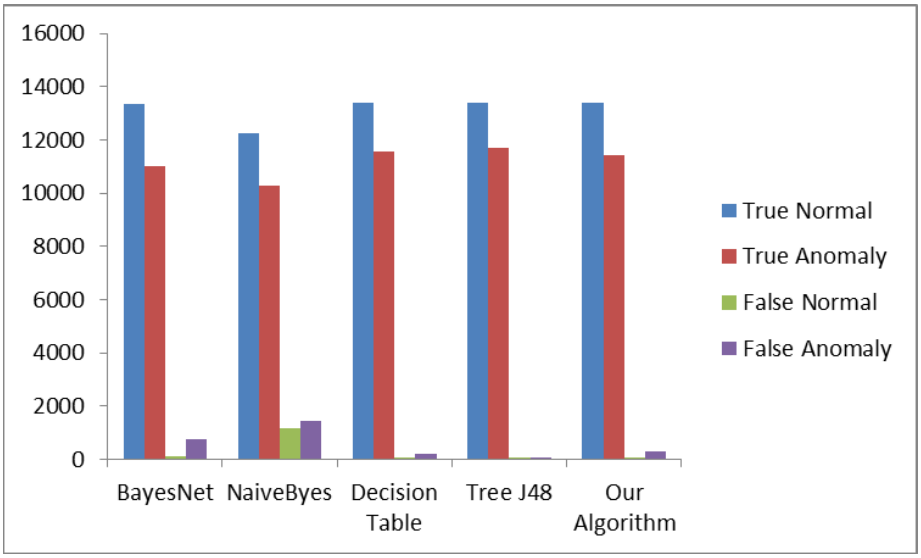

Fig. 2. A graphical representation of the confusion matrix.

Table 4. Comparative Classification Results and Time

Classifier	Correctly Classified Insistence	%	Incorrectly Classified Insistence	%	Time in ms
BayesNet	24326	96.5624	866	3.4378	0.63
NaiveByes	22570	89.5919	2622	2.8144	0.2
Decision Table	24925	98.9004	257	1.0202	11.55
Tree J48	25081	99.615	111	0.4406	1.55
Our Algorithm	24815	98.5035	377	1.4965	0.18

6. Conclusion

The paper proposes a novel approach that has better accuracy and time complexity. It helps to develop an application for multi-core, multi-threading inline intrusion detection and prevention system. It reduces the load to dedicated security devices in the network. It improves the overall performance by reducing the false attacks. It also takes very less time as compared to other classifiers as summarised in Table 4.

7. References

1. J. Lafferty, A. McCallum, and F. Pereira. “Conditional random fields: Probabilistic models for segmenting and labeling sequence data.” In Proceedings of Eighteenth International Conference on Machine Learning, ICML, pages 282– 289, 2001.
2. Y. Gu, A. McCallum, and D. Towsley, “Detecting Anomalies in Network Traffic Using Maximum Entropy Estimation,” Proc. Internet Measurement Conf. (IMC ’05), pp. 345-350, USENIX Assoc., 2005
3. Shah, H.; Undercoffer, J.; Joshi, A. "Fuzzy clustering for intrusion detection," Fuzzy Systems, 2003. FUZZ. The 12th IEEE International Conference on, vol.2, no., Pp. 1274- 1278 vol.2, 25-28 May 2003
4. Ye Du; Huiqiang Wang; Yonggang Pan, "A hidden Markov models-based anomaly intrusion detection method," Intelligent Control and Automation, 2004. WCICA 2004. Fifth World Congress on, vol.5, no., Pp. 4348- 4351 Vol.5, 15-19 June 2004.
5. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

6. A.N.M. Etesham Rafiq, M. Watheq El-Kharashi, and Fayez Gebali," A fast string search algorithm for deep packet classification," Elsevier, computer Communication 27 (2004) 1524-1538.
7. K. Salah, A.Kahtani, "Performance evaluation comparison of Snort NIDS under Linux and Windows Server", Elsevier, Journal of Network and Computer Application 33(2020)6-15
8. Qiang Wei, Yunzhao LI , Yang Chu, "A Predict DEterminstic Finite Automaton for deep Packet Inspection", Elsevier, Procedia engineering 29(2012)2156-2161
9. Milton L. Mueller, Handi Asghar, "Deep packet inspection and bandwidth management: Battles over BitTorrent in Canada and The United States," Elsevier, Telecommunications Policy 36(2012)462-475
10. MU Cheng, HUANG Xiaohong, WU Jun, MA Yan, "Network Traffic Signature Generation Mechanism Using Principal Component Analysis," China Communication, November 2013
11. N.Kannaiya Raja, Dr. K.Arulanandam, B.Raja Rajeswari, " TWO-LEVEL PACKET INSPECTION USING SEQUENTIAL DIFFERENTIATE METHOD", IEEE, 2012 International Conference on Advances in Computing and Communication, 978-0-7695-4723-7/12 $26.00
12. Uday Trivedi, "A Self-Learning Stateful Application Identification Method For Deep Packet Inspection,"IEEE.
13. Soheil Hassas Yeganeh, Molded Eftekhari, Yashar Ganjali, Ram Keralpura, Antonio Nucci, "CUTE: traffic Classification Using Terms, IEEE.
14. Shinya Kawano, Toru Okugawa, Taizo Yamamoto, Tomoharu Motono and Yasushi Takagi, "High-Speed DPI method using multi-stage packet flow analyses,"IEEE.
15. Junaid Shaikh, Markus Filler, Patrik Arlosand denis College, "Modeling and Analysis of Web Usage and Experience Based on Link-Level Measurements", 978-0-9836283-4-7© ITC
16. Gang He, Bo Sun, Yang Liu, Xiaochun Wu, "I-HASH MULTIPLE VARIOUS POSITION PATTERN MATCHING ALGORITHM On INTERNET APPLICATION IDENTIFICATION". 978-1-4673-2204-1/12/$31.00©2012 IEEE.
17. Tomasz Blow, Sara Ligaard Norgaard Hald, Tahir Riaz, Jens Myrup Pedersen, "A Method for Evaluation of Quality of Service in Computer Networks,". ICACT Transactions on Advanced Communication Technology (TACT) vol,1, issue 1, July 2012.
18. Long Zhong, Jinsong Wang, Sheng Lin, "Design of the Network Traffic Anomaly Detection System in Cloud Computing Environments," 2012 Fourth International Symposiums on Information Science and Engineering, 978-0-7695-4951-4/13 $26.00 © 2013 IEEE.
19. Han Zhang, Christos Papadopoulos, Dan Massey, "Detecting Encrypted Botnet Traffic," IEEE Global Internet Symposium.
20. Payam Mahdinia, Mehdi Berenjkoob, Hedayat VatankhH, "Attack Signature Matching using Graphics Processors I High-Performance Intrusion Detection Systems," IEEE, 978-1-4673-5634-3/3/13/$31.00 ©2013.
21. Y.Yang, K.McLaughlin, T.Littler, S.Sezer. "Intrusion Detection System for ICE 608070-5-104 Based SCADA Networks", 978-1-4799-1303-9/13 ©2013 IEEE.
22. Haiqiang Wang, Kuo-Kun Tseng and Jeng-Shenyang Pan, "A Novel Statistical Automation for Network Cloud Traffic Classification," 978-1-4673-2588-2/12 ©2012 IEEE
23. R.M. Karp, M.O. Rabin, "Efficient randomized pattern_matching algorithms, IBM Journal of Research and Development 31(2) (1987) 249-260.
24. Robert E. Crosser, Allen C. Johnston, Paul Benjamin Lowry, Quinn Hu, Merrill Warkentin, Richard Baskerville, "Future directions for behavioral information security research."
25. Roman Klinger, Katrin Tomanek, "Classical Probabilistic Models and Conditional Random Fields ", Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven
26. www.cavium.com/pdfFiles/CSS-DPI-White-Paper.pdf
27. www.ipoque.com/sites/default/../white-paper-deep-packet-inspection.pdf
28. www.datacom.cz/files_datacom/dpi_white_paper.pdf
29. Porto.polito.it/2375838/1/11JNSM_LightweigthDPI..