INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN COMPUTER SCIENCE AND ENGINEERING TECHNOLOGIES

ISSN : 2454-9924

NOVEL AUTOMATIC TEST PACKET GENERATION

K.KOLAPANENI OMSRI HARIKA[1], S.Suresh[2]

[1]M.Tech Student, Sree Rama institute of technology and science Kuppenakuntla,Penuballi,Khammam,TS INDIA

[2]Asst Prof,CSE DeptSree Rama institute of technology and science Kuppenakuntla,Penuballi,Khammam,TS INDIA

ABSTRACT:

. Networks are getting larger and more complex, yet administrators rely on rudimentary tools such as and to debug problems. We propose an automated and systematic approach for testing and debugging networks called “Automatic Test Packet Generation” (ATPG). ATPG reads router configurations and generates a device-independent model. The model is used to generate a minimum set of test packets to (minimally) exercise every link in the network or (maximally) exercise every rule in the network. Test packets are sent periodically, and detected failures trigger a separate mechanism to localize the fault. ATPG can detect both functional (e.g., incorrect firewall rule) and performance problems (e.g., congested queue). ATPG complements but goes beyond earlier work in static checking (which cannot detect liveness or performance faults) or fault localization (which only localize faults given liveness results). We describe our prototype ATPG implementation and results on two real-world data sets: Stanford University’s backbone network and Internet2. We find that a small number of test packets suffices to test all rules in these networks: For example, 4000 packets can cover all rules in Stanford backbone network, while 54 are enough to cover all links. Sending 4000 test packets 10 times per second consumes less than 1% of link capacity. ATPG code and the data sets are publicly available.

Index Terms-Data plane analysis, network troubleshooting, test packet generation.

INTRODUCTION:

I T IS notoriously hard to debug networks. Every day, network engineers wrestle with router misconfigurations, fiber cuts, faulty interfaces, mislabeled cables, software bugs, intermittent links, and a myriad other reasons that cause networks to misbehave or fail completely. Network engineers hunt down bugs using the most rudimentary tools (e.g., , , SNMP, and) and track down root causes using a combination of accrued wisdom and intuition. Debugging networks is only becoming harder as networks are getting bigger (modern data centers may contain 10 000 switches, a campus network may serve 50 000 users, a 100-Gb/s long-haufor machine learning applications. They can be divided into four broad categories: the Embedded, Wrapper, Filter, and Hybrid approaches The embedded methods incorporate feature selection as a part of the training process and are usually specific to given learning algorithms, and therefore may be more efficient than the other three categories . Traditional machine learning algorithms like decision trees or artificial neural networks are examples of embedded approaches. The wrapper methods use the predictive accuracy of a predetermined learning algorithm to determine the goodness of the selected subsets, the accuracy of the learning algorithms is usually high. However,

ISSN : 2454-9924

the generality of the selected features is limited and the computational complexity is large. The filter methods are independent of learning. algorithms, with good generality. Their computational complexity is low, but the accuracy of the learning algorithms is not guaranteed . The hybrid methods are a combination of filter and wrapper methods by using a filter method to reduce search space that will be considered by the subsequent wrapper. They mainly focus on combining filter and wrapper methods to achieve the best possible performance with a particular learning algorithm with similar time complexity of the filter methods. The wrapper methods are computationally expensive and tend to over fit on small training sets. The filter methods, in addition to their generality, are usually a good choice when the number of features is very large. Thus, we will focus on the filter method in this paper.

Existing System:

Inorder to getting a better predictor for that they provide mostly information which is already present in other feature(s). Of the many feature subset selection algorithms, some can effectively eliminate irrelevant features but fail to handle redundant features yet some of others can eliminate the irrelevant while taking care of the redundant features . Our proposed FAST algorithm falls into the second group. Traditionally feature subset selection research has focused on searching for relevant features. A well-known example is Relief which weighs each feature according to its ability to discriminate instances under different targets based on distance-based criteria function. However, Relief is ineffective at removing redundant features as two predictive but highly correlated features are likely both to be highly weighted. Relief extends Relief,

ISSN : 2454-9924

enabling this method to work with noisy and incomplete data sets and to deal with multiclass problems, but still cannot identify redundant features.

Proposed System:

Recently, hierarchical clustering has been adopted in word selection in the context of text classification Distributional clustering has been used to cluster words into groups based either on their participation in particular grammatical relations with other words by Pereira et al. or on the distribution of class labels associated with each word by Baker and McCallum . As distributional clustering of words are agglomerative in nature, and result in suboptimal word clusters and high computational cost, Dhillon et al. proposed a new Information theoretic divisive algorithm for word clustering and applied it to text classification. Butterworth et al. proposed to cluster features using a special metric of Barthelemy-Montjardet distance, and then makes use of the dendrogram of the resulting cluster hierarchy to choose the most relevant attributes. Unfortunately, the cluster evaluation measure based on Barthelemy-Montjardet distance does not identify a feature subset that allows the classifiers to improve their original performance accuracy. Further more, even compared with other feature selection methods, the obtained accuracy is lower.

FEATURE SUBSET SELECTION ALGORITHM

3.1 Framework and Definitions

Irrelevant features, along with redundant features, severely affect the accuracy of the learning machines .Thus, feature subset selection should be able to identify and remove as much of the irrelevant and redundant information as possible. Moreover, we develop a novel algorithm which

ISSN : 2454-9924

can efficiently and effectively deal with both irrelevant.

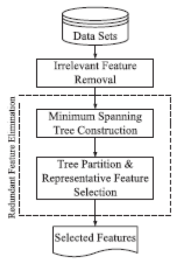
Data Sets
Irrelevant Feature Removal
Minimum Spanning Tree Construction
Tree Partition & Representative Feature Selection
Redundant Feature Elimination
Selected Features

Fig. 1. Framework of the proposed feature subset selection algorithm.

and redundant features, and obtain a good feature subset. We achieve this through a new feature selection framework (shown in Fig. 1) which composed of the two connected components of irrelevant feature removal and redundant feature elimination. The former obtains features relevant to the target concept by eliminating irrelevant ones, and the latter removes redundant features from relevant ones via choosing representatives from different feature clusters, and thus produces the final subset.

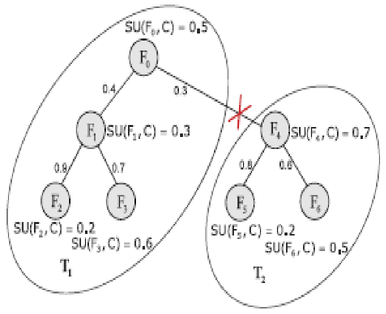

Fig. 2. Example of the clustering step.

Results and Analysis

In this section, we present the experimental results in terms of the proportion of selected features, the time to obtain the feature subset, the classification accuracy, and the Win/ Draw/Loss record. For the purpose of exploring the statistical significance of the results, we performed a nonparametric Friedman test followed by Nemenyi posthoc test as advised by Demsar and Garcia and Herrerato to statistically compare algorithms on multiple data sets. Thus, the Friedman and the Nemenyi test results are reported as well.

ISSN : 2454-9924

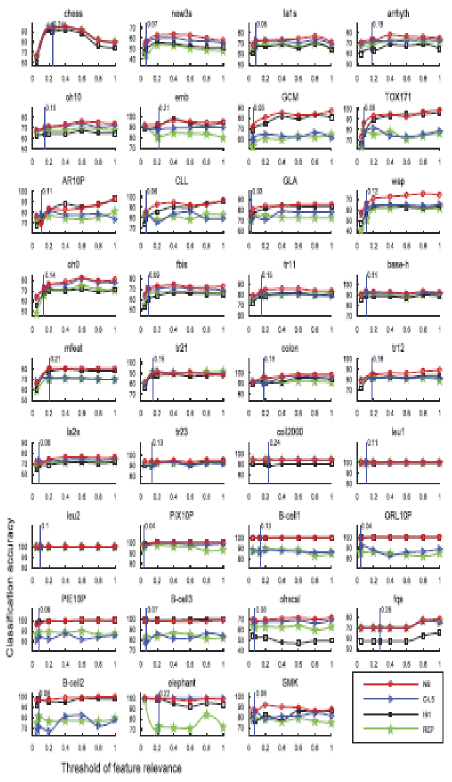

Fig. 9. Accuracies of the four classification algorithms with different θ values.

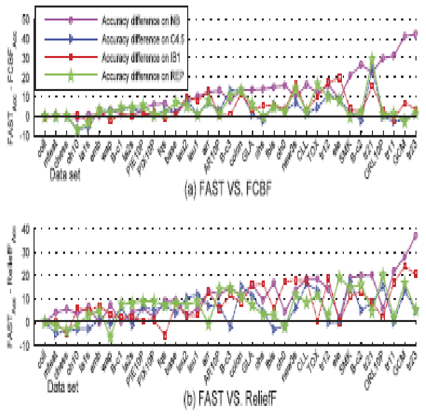

Fig. 10. Accuracy differences between FAST and the comparing algorithms.

CONCLUSION

Testing liveness of a network is a fundamental problem for ISPs and large data center operators. Sending probes between every pair of edge ports is neither exhaustive nor scalable [30]. It suffices to find a minimal set of end-to-end packets that traverse each link. However, doing this requires a way of abstracting across device specific configuration files (e.g., header space), generating headers and the links they reach (e.g., all-pairs reachability), and finally determining a minimum set of test packets ZENG et al.: AUTOMATIC TEST PACKET GENERATION 565 (Min-Set-Cover). Even the fundamental problem of automatically generating test packets for efficient liveness testing requires techniques akin to ATPG. ATPG, however, goes much further than liveness testing with the same framework. ATPG can test for reachability policy (by testing all rules including drop rules) and performance health (by associating performance measures such as latency and loss with test packets). Our implementation also augments testing with a simple fault localization scheme also constructed using the header space framework. As in software testing, the formal model helps maximize test coverage while minimizing test packets. Our results show that all forwarding rules in Stanford backbone or Internet2 can be exercised by a surprisingly small number of test packets (for Stanford, and for Internet2). Network managers today use primitive tools such as and . Our survey results indicate that they are eager for more sophisticated tools. Other fields of engineering indicate that these desires are not unreasonable: For example, both the ASIC and software design industries are buttressed by billion-dollar tool businesses that supply techniques for both static (e.g., design rule) and dynamic (e.g., timing) verification. In fact, many months after we built and named our system, we discovered to our surprise that ATPG was a well-known acronym in hardware chip testing, where it stands for Automatic Test Pattern Generation [2]. We hope network ATPG will be equally useful for automated dynamic testing of production networks.

ACKNOWLEDGMENTS

The authors would like to the editors and the anonymous reviewers for their insightful and helpful comments and suggestions, which resulted in substantial improvements to this work. This work is supported by the National Natural Science Foundation of China under grant 61070006.

REFERENCES

[1] H. Almuallim and T.G. Dietterich, "Algorithms for Identifying Relevant Features," Proc. Ninth Canadian Conf. Artificial Intelligence,

ISSN : 2454-9924

pp. 38-45, 1992.

[2] H. Almuallim and T.G. Dietterich, "Learning Boolean Concepts in the Presence of Many Irrelevant Features," Artificial Intelligence, vol. 69, nos. 1/2, pp. 279-305, 1994.

[3] A. Arauzo-Azofra, J.M. Benitez, and J.L. Castro, "A Feature Set Measure Based on Relief," Proc. Fifth Int'l Conf. Recent Advances in Soft Computing, pp. 104-109, 2004.

[4] L.D. Baker and A.K. McCallum, "Distributional Clustering of Words for Text Classification," Proc. 21st Ann. Int'l ACM SIGIR Conf. Research and Development in information Retrieval, pp. 96-103, 1998.

[5] R. Battiti, "Using Mutual Information for Selecting Features in Supervised Neural Net Learning," IEEE Trans. Neural Networks, vol. 5, no. 4, pp. 537-550, July 1994.

[6] D.A. Bell and H. Wang, "A Formalism for Relevance and Its Application in Feature Subset Selection," Machine Learning, vol. 41, no. 2, pp. 175-195, 2000.

[7] J. Biesiada and W. Duch, "Features Election for High-Dimensional data a Pearson Redundancy Based Filter," Advances in Soft Computing, vol. 45, pp. 242-249, 2008.

[8] R. Butterworth, G. Piatetsky-Shapiro, and D.A. Simovici, "On Feature Selection through Clustering," Proc. IEEE Fifth Int'l Conf. Data Mining, pp. 581-584, 2005.

[9] C. Cardie, "Using Decision Trees to Improve Case-Based Learning,"

Proc. 10th Int'l Conf. Machine Learning, pp. 25-32, 1993.

[10] P. Chanda, Y. Cho, A. Zhang, and M. Ramanathan, "Mining of Attribute Interactions Using Information Theoretic Metrics," Proc.
IEEE Int'l Conf. Data Mining Workshops, pp. 350-355, 2009.

[11] S. Chikhi and S. Benhammada, "ReliefMSS: A Variation on a Feature Ranking Relieff Algorithm," Int'l J. Business Intelligence and Data Mining, vol. 4, nos. 3/4, pp. 375-390, 2009.

[12] W. Cohen, "Fast Effective Rule Induction," Proc. 12th Int'l Conf. Machine Learning (ICML '95), pp. 115-123, 1995.

[13] M. Dash and H. Liu, "Feature Selection for Classification," Intelligent Data Analysis, vol. 1, no. 3, pp. 131-156, 1997.

[14] M. Dash, H. Liu, and H. Motoda, "Consistency Based Feature Selection," Proc. Fourth Pacific Asia Conf. Knowledge Discovery and
Data Mining, pp. 98-109, 2000.

[15] S. Das, "Filters, Wrappers and a Boosting-Based Hybrid for Feature Selection," Proc. 18th Int'l Conf. Machine Learning, pp. 74-81, 2001.

[16] M. Dash and H. Liu, "Consistency-Based Search in Feature Selection," Artificial Intelligence, vol. 151, nos. 1/2, pp. 155-176, 2003.

[17] J. Demsar, "Statistical Comparison of Classifiers over Multiple

ISSN : 2454-9924

Data Sets," J. Machine Learning Res., vol. 7, pp. 1-30, 2006.

[18] I.S. Dhillon, S. Mallela, and R. Kumar, "A Divisive Information Theoretic Feature Clustering Algorithm for Text Classification," J. Machine Learning Research, vol. 3, pp. 1265-1287, 2003.

K.KOLAPANENI OMSRI HARIKA is an M.Tech Depart ment of Co mputer Science & Engineering, Sreerama Institute of Technology & science, Penuballi Mandal, Khammam, Kotha Kuppenkuntla

S. Suresh well known author and excellent teacher. He is working as a H.O.D for the Department of M.Tech., Computer Science &
Engineering, Sree Rama Institute of Technology & Science, Kuppenakuntla, Penuballi, Khammam. He has vast teaching experience in various engineering colleges. To his
credit couple of publications both National& International conferences / journals. His area ofInterest includes Data Warehouse and Data Mining,information security, Data Communications &Networks, Software Engineering and otheradvances in Computer Applications. He has guided many projects for Engineering Students