ELECTRONICS AND ELECTRICAL ENGINEERING
ISSN 1392 – 1215 2008. No. 5(85)
ELEKTRONIKA IR ELEKTROTECHNIKA

T170 ELECTRONICS
ELEKTRONIKA

Data Mining for Managing Intrinsic Quality of Service in MPLS

J. Jeļinskis, G. Lauks
Institute of Telecommunication, Faculty of Electronics and Telecommunication, Riga Technical University
Āzenes str. 12, LV-51368 Riga, Latvia, phone: +371 7089201; e-mail: lauks@rsf.rtu.lv, jans.jelinskis@etf.rtu.lv

Introduction

Quality of Service (QoS) is more and more becoming a necessity for emerging applications carried by IP networks. Thus, dynamic admission control is a very important mechanism that aims, not only to ensure resources availability, but also to verify QoS constraints satisfaction [12]. Nowadays traffic engineering is an essential ingredient for guaranteeing QoS and for efficient or cost effective resource utilization, design and operation of IP networks. MPLS (Multiprotocol Label Switching) is a standards-based technology that can improve network performance and QoS for select traffic [2]. MPLS offers multiple classes of service, each associated with different locally significant labels in the packet header and the packets are forwarded by network nodes via label swapping types of traffic. MPLS uses short, fixed-length similar to layer 2 switching [3]. The resulting connections are termed label switched paths (LSP) and a router that supports the MPLS protocols is called a label switching router (LSR). However LSP set up admission control policy is one of the notable problems that have to be solved to fulfill the requirements for effective resource allocation and network utilization for appropriate QoS level.

The off-line network design methods, which use hypothetical knowledge of traffic demand, are not suitable for MPLS networks [8] because of the high unpredictability of the internet traffic. A fully connected MPLS network, where direct LSPs link every pair of LSRs, is very inefficient [9] because of very high signaling cost and the management of a huge number of LSPs. The signaling cost is of the order of N^2, where N is the total number of routers.

Simple LSP setup policies, based on traffic driven approach, in which an LSP is established whenever the number of bytes forwarded within specified time boundaries exceeds a threshold, are not suitable because of very high signaling costs and high control efforts for variable and bursty traffic [10]. There is need of new traffic-driven approaches that are usable with variable and bursty traffic requests, because internet traffic has been found to exhibit significant amounts of self similarity and long range dependence [11] do to an extremely high variability of burst duration.

Knowledge discovery in databases (KDD) is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns from large data collections [1]. Data Mining now refers to one stage within the KDD process for extracting useful rules and patterns from the data.

Data mining is a logical process that is used to search through large amounts of information in order to find important data. The goal of this technique is to find patterns that were previously unknown [4]. Once you have found these patterns, you can use them to solve a number of problems.

This paper sketches the reference scenario of Data Mining approach in managing LSP set up policy in MPLS networks and discusses the subject of self-adapting approach, attempting to focus on the key issues of Decision Tree algorithm performance. As a relevant example, this paper briefly reports an algorithm of Data Mining approach to LSP set up problem developed by the authors that implements a traffic engineering solution able to fulfill the contemporary requirements. This approach could be regarded as a one step toward the realization of self-adapting networks. In addition, some relevant results that were obtained by simulation are reported to discuss the main characteristics of such a system and assess the feasibility of the concept. Finally, the paper reviews the main hot issues that need to be discussed, in the authors' opinion, in the future researches.

Performances of Decision Trees

Decision trees are powerful and popular tools for classification and prediction. The attractiveness of decision trees is due to the fact that, in contrast to neural networks, decision trees represent rules. Rules can readily be expressed so that humans can understand them or even directly use in a database access language like Structured Query Language (SQL) so that records falling into a particular category may be retrieved [6].

There are a variety of algorithms for building decision trees that share the desirable quality of interpretability. In this research we have used a well known and frequently used over the years C4.5 algorithms, which is an extension of ID3 that accounts for unavailable values, continuous attribute value ranges, pruning of decision trees, rule derivation, and so on [7].

The Decision Tree algorithms produce classification and regression models in a tree-based structure. In this paper, we consider only the DT classification property, that is, the mapping of an input vector (observation) x, that can be represented as (1) to a class label (decision).

At each node, the DT algorithm searches through the variables one by one, and for each variable it finds the best split according to the split criterion. Finally, the algorithm compares the n best single variable splits and selects the best of the best splits [4].

Classification with the DT is based on observations of a set of variables data, variables which are used as predictors, and a classification variable attached to these observations.

The problem to be solved in this tree construction is to determine the binary splits of data set X with training data set L so that X is cut into smaller and smaller subsets [4]. Algorithms searches over every possible threshold in every variable for the split that best improves the tree structure according to a specified score function

Classification trees offer an effective implementation of such hierarchical classifiers. Indeed, classification trees have become increasingly important due to their conceptual simplicity and computational efficiency [5].

LSP setup problem

When a bandwidth request arrives between two nodes in a network that are not connected by a direct LSP, the decision about to establish such a LSP arises.

Decision of establishing LSP is based on lots of arguments, who can be presented as vector (1)

$$D_a = (a_1, a_2, a_3 \ldots a_k). \quad (1)$$

There is no one strict state of argument vector D_a, which for certain define objective of setting up LSP. In this case large number of objectives is suitable [13], and to find them, an Evolutionary Multiobjective Search is proposed with Genetic Algorithm.

Multiobjective optimization problem can be expressed as:

$$\text{minimize } f(x)=(f1(x), \ldots, fk(x)); \quad \text{subject to } x \in X ; \quad (2)$$

where x represents a solution, and X is a finite set of feasible solutions. We can`t use term “minimize” just so, because in general, there does not exist a single solution, that is minimal to all objectives. That means, we can try to find a set of solutions $X^* \subseteq X$, called the Pareto optimal set, with the property that:

$$\forall x^* \in X^*, \exists\ x \in X \text{ such that } x > x^*, \quad (3)$$

where $x > x^*$ if $\forall i \in \{1, \ldots, k\}$,

$f(x_i) \leq f(x_i^*) \ \wedge \ \exists i \in \{1, \ldots, k\} : f(x_i) < f(x_i^*)$; $x > x^*$ is read as x dominates x*, and solutions in the Pareto optimal set are also known as admissible solutions [14].

GA with following pseudocode was used:

a) Choose initial population randomly;
b) Evaluate the fitness function of each individual in the given population;
c) Do…
- Select best-ranking individuals to reproduce;
- Breed new generation through crossover and mutation (genetic operations) - make offspring;
- Evaluate the individual fitness of the new offspring;
- Replace worst ranked part of given population with offspring;

Until… <terminating condition>.

For chromosome modeling such arguments were used as utilization and losses, as functions of available bandwidth and requested traffic.

For traffic generation Glen Kramer self similar traffic generator [15] was used. In this generator, the resulting self-similar traffic is obtained by aggregating multiple sub-streams, each consisting of alternating Pareto-distributed on/off periods.

The load generated by one sub-stream is measured as $\lambda = E[on]/(E[on]+E[off])$, where $E[on]$ and $E[off]$ are expected lengths of on and off periods respectively [15]. The total load generated by the traffic generator is equal the sum of loads generated by all sub-streams.

As a result Pareto front of suitable states of network parameters was found, which are shown in Fig. 1.

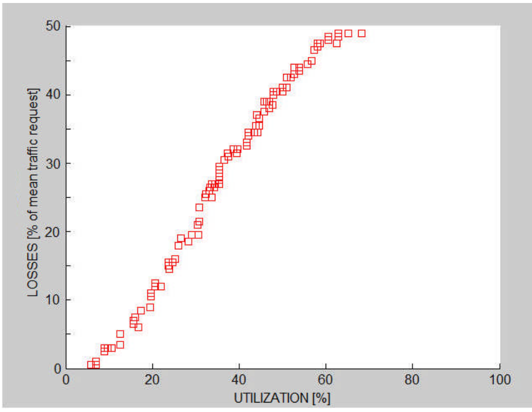

Fig. 1. Pareto front of admissible network states

Application of algorithm

In our recent paper [3] we have verified a possibility of a new LSP setup policy algorithm, which uses the same optimization procedure based on multi-objective model with Pareto ranking and Genetic Algorithm. We have assumed that adaptation to variable and bursty traffic is achieved by using Neural Network learning capabilities, and decision making under uncertainty is made up with Fuzzy Logical approach. However this approach has lots of stages and requires extremely precise optimal training data set for Neural Network and optimal ratio between learning and operating phases as well as active time of both phases, depending on amplitude changes of Self-similar traffic in time scale [3]. As the next step in developing LSP set up policy algorithm we propose to simplify it by using Data Mining technique – by applying Decision Tree (DT) approach to the selected QoS Key Performance Indicators (KPIs) for classification, instead of using Neural Network learning capabilities.

In allocating resources, we simplified the resource management system to LSP set-up policy between LSR. The search space is spanned by quality related key performance indicators (KPIs), generated to the database. In this research we limit numbers of possible KPIs to link utilization U_l, losses L, and bandwidth request R_{LSP}. Quantity of parameters, that represents QoS, is purposely reduced to simplify problem. In fact, given approach can be adjusted to one's liking number of parameters allowing for calculation time and measure of precision.

LSP setup admission algorithm, which is presented in this paper provide following classification and operating manner.

All the activities are divided into classification phase and operating phase. The first time, only classification phase is active, but after operating phase activation, both, classification and operating phases are present

a) Classification phase.

Statistical data are collected persistently; based on them, with Genetic Algorithm means the Pareto admissible solutions are found. Data file with admissible and synthesized non admissible network states is generated. Classification is performed.

a) Operating phase.

Operating phase is performed together with classification phase, which in this case is performed in background, to provide data for classification. In operating phase, network parameters and traffic request are obtained directly from network, and rules, generated by decision tree after classification allows to make direct decision – mainly whether establish LSP or not.

Classification and operating phase lengths or ratio of these two phases was not subject of this research, and this issue is supposed to be objective of future research.

In brief, algorithm proposed in this paper can be depictured as follows in Fig. 2.

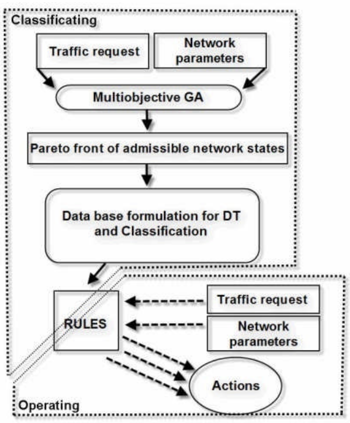

Fig. 2. LSP setup admission control algorithm

Feasibility of the concept

To validate the correctness of proposed approach simulation of algorithm was made. Classification process in this research was simplified to the binary classificator – True or False, that describes LSP set up possibility, and classification results obtained from used algorithm are depicted on Fig.3

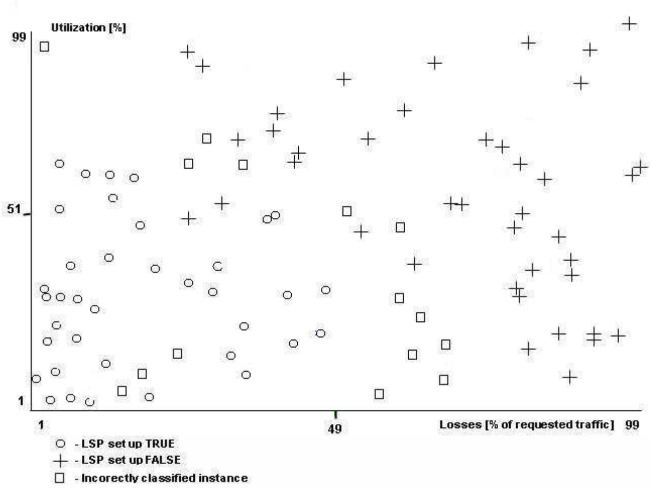

Fig. 3. LSP setup data classified instances and incorrectly classified observations

It is vitally important, that final decision making error is not exceeding certain determined threshold. It is depending on DT classification error, which is in its turn depending on number of observation, outliers in the classified data and many other factors. In our simulation we have used different data compositions with synthesized outliers to evaluate possible implications – see Fig.4.

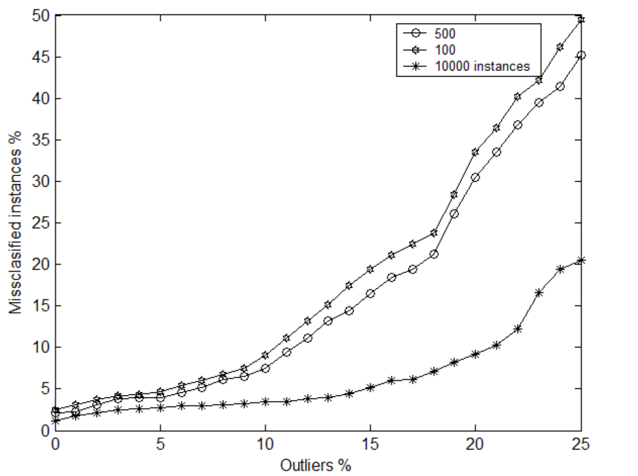

Fig. 4. Misclassification dependence of outliers in the inspected data

Conclusion and future research

In this paper we verify a possibility of LSP setup admission control algorithm, which uses optimization procedure based on multi-objective model with Pareto ranking and Genetic Algorithm. Using Data Mining classification capabilities algorithm allows making a decision under uncertainty with a certain error rate.

To ensure the contracted QoS is sustained, network must monitor QoS parameters and reallocate resources in response to system anomalies in on-line regime. This requires monitoring resource availability and its dynamic characteristics to detect deviations in the QoS parameters.

Data warehouses containing huge amounts of data can be made of monitoring measurement traces, and data analysis can be performed to make suitable decisions. This global approach can be applied to overall MPLS network monitoring and configuring, and it is the subject of future research.

References

1. **Usama M. Fayyad, Ramasamy Uthurusamy.** Data Mining and Knowledge Discovery in Databases // Communications of the ACM. – 1996. No. 39(11). – P. 24–26.
2. **Girish M., Bei Zhou, Jian-Qiang Hu.** An algorithm for rerouting in traffic engineering of MPLS based IP networks // IP Operations and Management. 3rd IEEE Workshop. – 2003. P. 115–118.
3. **Lauks G., Jeļinskis J.** Optimal Policy for LSP Control in MPLS Networks // Electronics and Electrical Engineering. – Kaunas: Technologija, 2007. – No. 5(77). – P. 65–68.
4. **Pekko Vehviläinen.** Data Mining for Managing Intrinsic Quality of Service in Digital Mobile Telecommunications Networks // PhD Thesis. – Tampere University of Technology. – 2004. – Publications 458.
5. **Mahesh Pal.** Ensemble Learning with Decision Tree for Remote Sensing Classification // Proceedings of World Academy of Science, Engineering and Technology. – 2007. – Vol. 26. – P. 735–737.
6. **Govindarajan M.** Text Mining Technique for Data Mining Application // Proceedings of World Academy of Science, Engineering and Technology. – 2007. – Vol 26. – P. 544–549.
7. **Weida Wang, Wenjian Liu.** A Hybrid Inductive Learning-based Scheduling Knowledge Acquisition Algorithm // Computational Intelligence for Modelling, Control and Automation. – Nov. 2006. – P. 1.
8. **Kodialam M., Lakshman T. V**. Minimum interference routing with applications to MPLS traffic engineering // Proceedings of IEEE INFOCOM. – Tel Aviv, Israel. – 2000. – Vol. 2. – P. 884–893.
9. **Uhlig S., Bonaventure O.** On the cost of using MPLS for interdomain traffic // Quality of Future Internet Services. – Berlin, Germany. – 2000. – P. 141–152.
10. **Anjali T., de Oliveira J. C., Akyildiz I. F., Uhl G.** Optimal policy for label switched path in MPLS networks // Computer Networks 39. – 2002. – P. 165–183.
11. **Yousefi`zadeh H.** Neural Network Modeling of a Class of on-off Source Models with Self-Similar Characteristics // Proc. Of the First Workshop on Fractals and Self-Similarity, ACM SIGKDD. – 2002. – P. 419–421.
12. **Hichem Ayed Harhira, Samuel Pierre.** Multi-path dynamic admission control in MPLS networks with end-to-end delay guarantees // Proceedings of the 2nd ACM workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks. – Chania, Crete Island, Greece. – 2007. – P. 45–49.
13. **Sanchez G., Gomez-Skarmeta A. F., Roubos H., Babuska R.** Fuzzy Modeling with Multi-Objective Neuro-Evolutionary Algorithms // Systems, Man and Cybernetics, IEEE International Conference. – October 2002. – Vol.3. – P. 6–9.
14. **Joshua Knowles, Martin Oates, David Corne.** Advanced Multi-objective Evolutionary Algorithms Applied to two Problems in Telecommunications // BT Technology Journal. – October 2000. – No. 18(4). – P. 51–65.
15. **Glen Kramer.** Generator of Self-Similar Network Traffic // 2043 John D. Kemper Hall, University of California. Available at: www.csif.cs.ucdavis.edu/~kramer/code /traffic.

Submitted for publication 2008 02 15

J. Jeļinskis, G. Lauks. Data Mining for Managing Intrinsic Quality of Service in MPLS // Electronics and Electrical Engineering. – Kaunas: Technologija, 2008 – No. 5(85). – P . 33–36.

LSP set up admission control policy is one of the notable problems that have to be solved to fulfill the requirements for effective resource allocation and network utilization for appropriate QoS level. In this paper, we verify a possibility of a new LSP setup admission algorithm, which uses optimization procedure based on multi-objective model with Pareto ranking and Genetic Algorithm. Decision rules are generated with Data Mining approach by performing classification operation to the selected data. This algorithm functions in two phases – classification and operating, which are accomplished consecutive. Algorithm is described and depicted. Experimental data are depicted and future research subjects are pointed. Ill. 4, bibl. 15 (in English; summaries in English, Russian and Lithuanian).

И. Иелинскис, Г. Лаукс. Использование данных MPLS для управления качества // Электроника и электротехника. – Каунас: Технология, 2008. – № 5(85) – С. 33–36.

Описываются основы управления уровня качества сетей с целью эффективного распределения ресурсов. Предложен алгоритм, когда этап классификации оканчивается этапом управления, основанным на генетических алгоритмах. Даны экспериментальные результаты и направления исследований в будущем. Ил. 4, библ. 15 (на английском языке; рефераты на английском, русском и литовском яз.).

J. Jeļinskis, G. Lauks. Duomenų gavybos panaudojimas MPLS kokybei valdyti // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2008. – Nr. 5(85). – P. 33–36.

LSP prieigos valdymo politika yra viena esminių problemų, kurią reikia išspręsti norint patenkinti efektyvaus išteklių paskirstymo reikalavimus ir užtikrinti tinkamą tinklo eksploatavimo kokybės lygį. Darbe analizuojamos naujos LSP prieigos politikos galimybės. Joje naudojama daugiatiksliu modeliu pagrįsta optimizavimo procedūra, apimanti Pareto ir genetinius algoritmus. Sprendimo priėmimo taisyklės generuojamos naudojant duomenų gavybos (angl. Data Mining) metodą, klasifikuojant surinktus duomenis. Algoritmo veikimas susideda iš dviejų vienas po kito einančių etapų – klasifikavimo ir vykdymo. Aprašomas algoritmas, pateikiami eksperimentiniai duomenys bei siūlomos ateities tyrimų kryptys. Il. 4, bibl. 15 (anglų kalba; santraukos anglų, rusų ir lietuvių k.).