On the efficiency of multi-domain routing

ZOLTÁN CSERNÁTONY

Budapest University of Technology and Economics,
Department of Telecommunications and Media Informatics
csernatony@tmit.bme.hu

Keywords: multi-domain, aggregation, routing, scalability, efficiency

In order to cope with scalability issues, large networks are often divided into several domains where routing is performed only on the basis of their aggregating structures. Although scalability can be assured this way, further inaccuracies are introduced in the routing information affecting eventually the efficiency of routing. In our paper, a novel statistical threshold-based aggregation model is suggested with which we compare existing routing methods on the basis of the efficiency they provide. The method itself is a statistical extension of the standard poly-line segment method where re-aggregation is triggered on the basis of measurement data. According to the simulation results, our statistical method outperforms existing methods.

1. Introduction

In large communication networks, scalability issues occur as (1) traditional routing algorithms perform inefficiently in the presence of large numbers of nodes, and (2) distributing a large amount of routing advertisement might be infeasible due to the concerning instabilities [1] and convergence criteria of certain routing protocols [2].

In order to assure a scalable networking operation, nodes are grouped into smaller *domains* (sub-networks), and the topologies of these domains are *aggregated* to simpler structures with domains only distributing (routing) advertisements concerning only their aggregating representation. By applying such a technique, a scalable *multi-domain* network can be obtained which is actually a hierarchical structure consisting of a lower layer meant by the original network with an upper layer of the aggregating topologies[1].

Although topology aggregation assures scalability, it also has an unpleasant side-effect as it introduces certain inaccuracies in the routing advertisements[2]. Eventually, these inaccuracies might result in inefficient path selections due to the loss of information which the routing itself is executed on. As the premises suggest, a general property of applying topology aggregation is that the amount of advertisement information gets traded with *routing efficiency*.

Although there have been important researches done on topology aggregation and its effects on multi-domain routing, the direct relationship of scalability and routing efficiency is yet to be examined. As routing efficiency is related to the accuracy of aggregating information, contributions relevant to our concentration have been already made. An earlier study shows that, given the inaccuracy of aggregating representations, finding paths that are likely to satisfy requirements for parameters of additive metrics is NP-hard [3]. For a directed graph, [4] presents a compact aggregating representation that bounds the aggregation-induced inaccuracies by a worst-case distortion factor in case of a single additive parameter. Either for a single additive or for a restrictive parameter, [12] presents a distortion-free representation applying full mesh and spanning tree, respectively.

Although [5] demonstrates that even though topology aggregation reduces the routing information to a large extent, it does not necessarily diminish routing efficiency as considerably. In [6] and [15], a line-segment approach is introduced for topology representation by which the possible bandwidth and delay values are represented by a single linear approximation. Further improvements to this method apply estimations by poly-lines and (higher-degree) curves instead of single lines [7,8], among which poly-lines proved to be the most efficient. Another approach of aggregation captures the statistical properties of the original network by histograms with probabilities in order to handle uncertain parameters [9]. Study [10] employs a spanning tree method by which the topology is aggregated incrementally according to the histogram consisting of measurement data, while [11] applies partial link-based advertisement which is controlled by monitoring only the portion of the physical with the largest contribution to link-state variation.

In this paper, we suggest a novel threshold-based statistical aggregation model which is actually a measurement-driven extension of the poly-line segment method which has proven to be a highly efficient method [8]. For the purpose of evaluation, scalability and routing efficiency are evaluated in the case of four different aggregation methods. On the basis of the efficiencies achievable by them, we make comparisons by means of simulation.

1 Due to the simpler topology (i.e. sharing less sensitive information), an increased level of confidentiality is an additional benefit of aggregation.
2 Although the stochastic nature of traffic introduces further inaccuracies, that particular aspect is out of our study's scope.

The rest of the paper is organized as follows. In Section 2, we present the multi-domain network model, the aggregation method and the routing model, while Section 3 introduces functions for evaluating the scalability and routing efficiency. In Section 4, we evaluate these efficiency quantities by means of simulation, and finally, Section 5 summarizes the paper.

2. Topology aggregation and routing model

Aggregation models provide more accurate aggregating representation for a larger amount of information. Consequently, scalability is expected to be traded with routing efficiency. In this section, we give a formal foundation of the applied aggregation and routing models.

2.1 Modeling multi-domain networks

In this paper, we use similar basic assumptions in the network modeling as there have been in [6]. A multi-domain network is regarded as several interconnected domains where the terminating nodes of an interconnecting edge are called *border nodes*. We represent a domain's graph by a tuple *(V,B,E)* where *V* denotes the set of nodes, *B* the set of border nodes ($B \subseteq V$) while *E* denotes the set of edges.

Two traffic attributes are considered in this paper: *bandwidth* and *delay*, which are of *restrictive* and *additive* metrics, respectively. The operation by which the resultant value of a path *p* is to be computed determines the metrics of attributes, which are in our case

$$b[p] = \min_{e \in p} b[e]$$
$$d[p] = \sum_{e \in p} d[e]$$

where $b[e]$ and $b[p]$ denote the bandwidth of edge *e* and path *p*, respectively, and we use *d* for denoting delay. *Fig. 1* illustrates the operations according to *restrictive* and *additive* metrics.

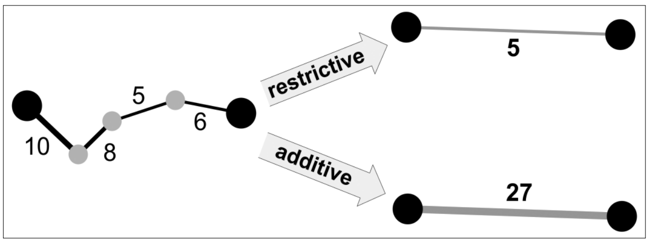


Figure 1. Resultant path values for restrictive (bandwidth) and additive (delay) attributes

2.2 Aggregation models

In this section, on the basis of their resulted efficiencies, we compare our *Statistical Threshold-based Model* (STM) with the classical *Poly-line Segment Method* (PLSM) for topology aggregation and the rather extreme *Best point* (BP) and *Worst point* (WP) approaches. These models provide aggregating topologies which are simplified abstractions of the physical topologies as they contain only the border nodes with aggregating edges ("logical links") running between them $(V,B,E) \Rightarrow (B,E_{aggr})$. This way, a two-layer network representation can be obtained as being illustrated in *Fig. 2*.

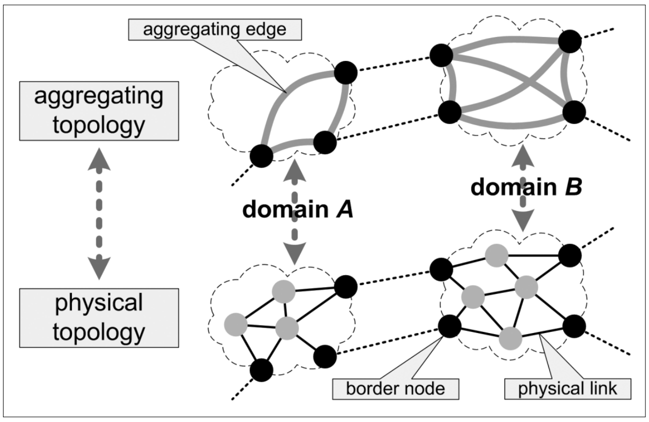


Fig. 2. Topology aggregation of domains: an illustration (two-layer model)

Poly-Line Segment Method (PLSM): This method [7,8] estimates the delay-bandwidth function by fitting (by least squares method) line segments onto such points on the bandwidth-delay plane that consist of the minimum delays for all possible bandwidth values. These points are called *representatives* and they form a staircase on the bandwidth-delay plane [8]. The bandwidth axis is then divided into *L* disjoint intervals and a single line is fit to the representatives within each of these intervals.

Best Point (BP) and Worst Point (WP) approaches: These are used for benchmarking purposes as they mean the two extremities of aggregation policies: BP uses the maximum bandwidth and the minimum delay values, while WP uses the maximum bandwidth and the maximum delay of all representatives [6,8].

Statistical Threshold-based Model (STM): Contrarily to PLSM, it follows a different concept as – instead of the representatives – the line segments are fit onto the *average* (measured or estimated) *delay values,* triggered by predefined *(aggregation) thresholds*. The model addresses bandwidth intervals

$$[0,l_L],\ [l_1,l_L],\ [l_2,l_L],\ \ldots,\ [l_{L-1},l_L]$$

by *(bandwidth) levels*

$$0 < l_1 < l_2 < \ldots < l_L$$

The resulting aggregating topologies are denoted by

$$(B,E_1),\ (B,E_2),\ \ldots,\ (B,E_L)$$

where in order to support connection requests with the maximum possible bandwidth requirements (i.e. the maximum bandwidth link capacity), $l_L = \max\{b(e) : \forall e \in E\}$ should hold.

More detailed (thus more accurate) representation can be made by involving more levels in the aggregation. In case of a certain pair of border nodes, for interval $\lfloor l_j, l_{j+1} \rfloor$, the aggregation method consists of two steps:

1) *Determining topology* (B,E_j):

All the edges with insufficient bandwidth values get filtered out according to criterion $b(e) < l_j\ (\forall e \in E)$. There is an aggregating edge between two border nodes only if there exists a path between them after the edge-filtering.

2) *Determining the linear delay-bandwidth function:* A linear function is fit (by least squares method) on the actual delay values of all the connection requests with bandwidth demand $l_j \le b \le l_{j+1}$ that have traversed them[3]. Each domain is assumed to store the actual (outcome) bandwidth request-delay data points of previous connection requests for each pair of border nodes in a simple First-In First-Out (FIFO) storage of length K_{FIFO}. Aggregation (line fitting) is performed once the number of successful connection requests exceeds predefined *threshold* K_{aggr} so that

$$k \bmod K_{aggr} = 0$$

is satisfied. The initial data points are determined by PLSM (i.e. based on static values).

In *Fig. 3*, STM is illustrated. Note that bandwidth is represented through the intervals, implicitly.

2.3 Routing model

Bandwidth and delay are taken into consideration separately by the following steps:

1) *Composing the aggregating view of the network:* Selecting the aggregating topology for each domain that corresponds to the minimum bandwidth level in accordance with the current bandwidth request: $\min\{j: l_j \ge r, j=1,...,L\} \Rightarrow (B,E_j)$. In case of BP and WP, L=1 holds formally.
2) *Logical routing:* Routing on the basis of the network consisting of the selected aggregating topologies. Dijkstra's algorithm is performed on the graph according to the selected level with its corresponding aggregating (advertised) edge delay values.
3) *Actual (physical) routing:* Routing by Dijkstra's algorithm by each selected domain on the basis of their physical topologies and link delay values between its selected pair of border nodes.

As a feedback, each domain is assumed to get notified of the outcome values of all the transfers that have been initiated by it.

3. The efficiency of multi-domain routing

In order to compare different aggregation schemes, there should be functions defined for "measuring" the *efficiency* they provide. This section applies two concepts for efficiency evaluation: the first one aims to measure the information saving due to aggregation, while the second one is to express the optimality of path selections. Section 3.1 and 3.2 discuss the former and latter ones, respectively.

3.1 Aggregation efficiency

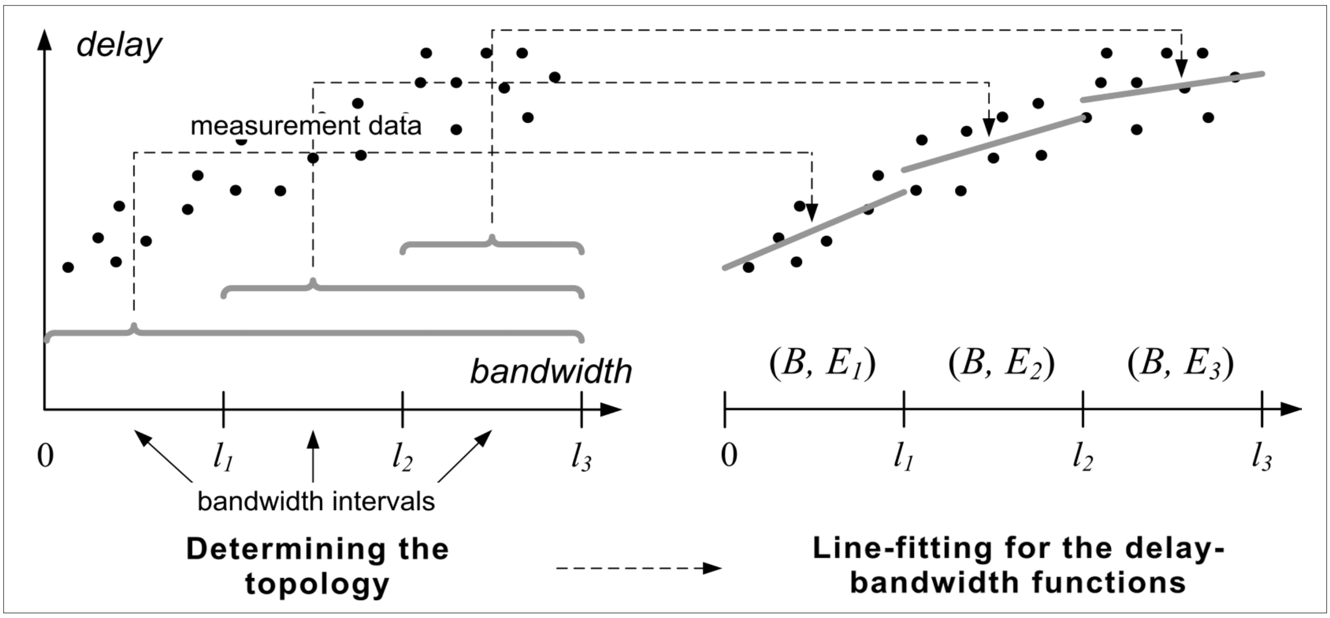
The amount of information to be advertised is closely related to scalability. Upon this fact, it is important to represent the (relative) amount of advertisement information that can be saved by applying topology aggregation. We use the following definition for expressing *aggregation efficiency*:

$$\eta_{AGGR} = \frac{I_{orig} - I_{aggr}}{I_{orig}} \tag{1}$$

where I_{orig} and I_{aggr} denote the amount of information meant by the original and aggregating topologies, respectively. Since a graph can be described by its edge set, assuming that each of the two attributes means a unit of information per edge, the following expressions hold: $I_{orig} = 2 \cdot |E|$ and $I_{aggr} = \sum_{j=1}^{L} |E_j|$. By evaluating them into (1), we obtain

$$\eta_{AGGR} = 1 - \frac{I_{aggr}}{I_{orig}} = 1 - \frac{1}{2}\frac{\sum_{j=1}^{L}|E_j|}{|E|} \tag{2}$$

Fig. 3. Statistical approximation via the threshold-based model by fitting line segments on measurement data for different bandwidth intervals (the case of three bandwidth intervals)



3 *Consequently, this model requires the deployment of such a monitoring method along with a database which keeps track of the observed traffic data. In this paper, however, we do not consider the realization of such requirements.*

Note that, on the one hand, due to inequality $I_{aggr} \geq 0$, upper bound $\eta_{AGGR} \leq 1$ holds. Also note that by applying refined aggregation models (e.g. star topology instead of meshes), the aggregation gain can be further increased, however this issue is out of our study's scope. Note that this result applies to both PLSM and STM since they only differ in the input data the aggregations are executed on.

3.2 Routing efficiency

Besides scalability, another important aspect is to evaluate the efficiency of routing. For that purpose, we introduce a *routing efficiency* function for both bandwidth and delay. We define these functions to be of normalized values for the range [0,1] and to be measureable by a domain on the basis of advertised and outcome values.

In the case of bandwidth (i.e. bandwidth sufficiency), we define *success ratio* as:

$$\eta_{succ} = \frac{total\ number\ of\ accepted\ requests}{total\ number\ of\ requests} \tag{3}$$

In the case of delay – as an inverse quantity of delay deviation [9] – we consider *delay efficiency* as the distance of the outcome (i.e. actual) values from the advertised values for the selected paths:

$$\eta_d = 1 - \frac{1}{N}\sum_{n=1}^{N}\left|a_{out}(n) - a_{adv}(n)\right| \tag{4}$$

where $a_{adv}(n)$ and $a_{out}(n)$ denote the advertised and outcome summarized values along the path selected at the n-th routing occasion, respectively. Since both $a_{adv}(n)$ and $a_{out}(n)$ are normalized for interval [0,1], bounds $0 \leq \eta_d \leq 1$ hold as well.

4. Simulation studies: Evaluating aggregation and routing efficiencies

In this section, we examine scalability and routing efficiency in case of PLSM and STM by evaluating the efficiency functions presented in Section 3.1 and 3.2. For the simulation testbed, we apply a combination of real-life approximating generated networks.

4.1 Simulation testbed

The simulation *testbed* is a two-layer multi-domain network whose inter-domain topology is generated by the *power-law model* [13] and the intra-domain topologies are generated based on the *Waxman model* [14]. The power model determines the number of degrees of a domain (i.e. the number of connected inter-domain links) according to formula

$$P(k) = c \cdot k^{-y} \tag{5}$$

where $P(k)$ denotes the probability of a node having k degrees and c is a normalizing constant that ensures $\sum_{k=1}^{D-1} P(k) = 1$, where D is the number of domains (there is no domains with zero degree (i.e. isolated domains)). On the other hand, Waxman model determines the probability that nodes u and v are connected by a physical link:

$$P(u,v) = b \cdot e^{-\frac{d(u,v)}{\beta L}}$$
$$L = \min_{\forall u,v\ u \neq v} d(u,v) \tag{6}$$

where the (physical) distance of nodes u and v is denoted by $d(u,v)$, while L denotes the minimum distance of any two nodes and b is a constant for normalization. The network consists of 20 domains with 30 nodes per domain. The parameters are set as been in [6], where parameters $\beta = 0.6$ and $y = 2.2$ are applied. Values from the normalized interval [0,1] are randomly assigned to each edge as bandwidth capacity (i.e. max. bandwidth) and delay: $b(e), d(e) \in [0,1]$, $\forall e \in E$.

In this paper, we apply a uniform distribution for the bandwidth demands of connection requests on the interval [0,1]. A dynamic traffic shape model is employed in which arriving times and holding times are distributed

Fig. 4. Bandwidth intervals vs. success ratio and aggregation efficiency

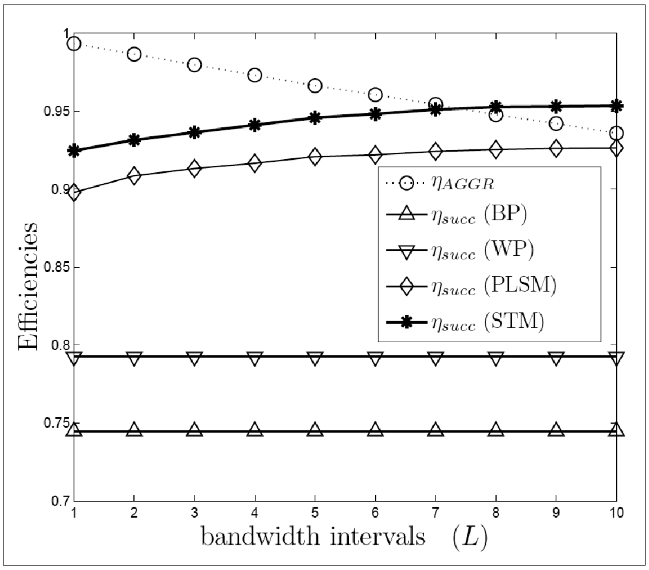
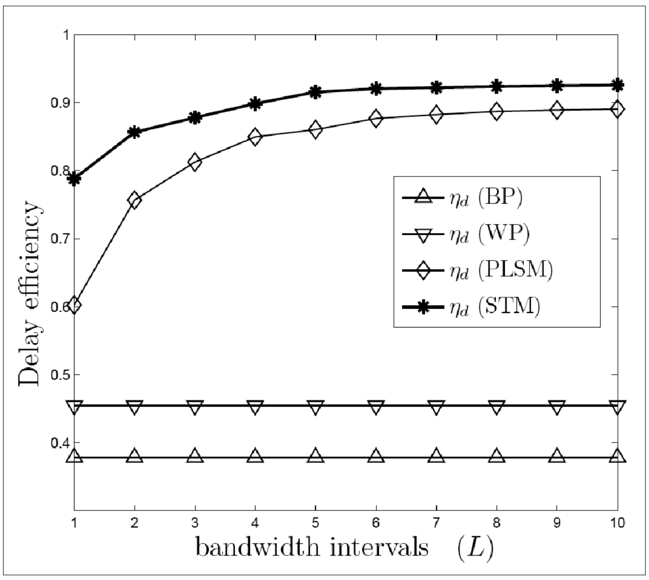


Fig. 5. Bandwidth intervals vs. delay efficiency



uniformly. For each number of bandwidth intervals and aggregation schemes, 1000 connection requests have been generated and averaged. In case of the two aggregation models PLSM and STM, for the sake of simplicity, the bandwidth levels (thus the intervals themselves) are distributed uniformly on the bandwidth axis. We use one up to ten different bandwidth levels $L=0,1,...,10$. For STM, the value of the aggregation threshold is set to $L\cdot 20$ and FIFO storage sizes (see Section 2.2) are set to $L\cdot 10$.

4.2 Simulation results

Our simulation results for BP, WP, PLSM and STM are presented in *Fig. 4* and *5*. Fig. 4 shows the evaluation of aggregation efficiency and success ratio for the four different aggregation schemes, while Fig. 5 shows the corresponding delay efficiencies depending on the number of bandwidth intervals. All efficiency values are computed as the average of efficiencies achieved by all domains. Note that in case of BP and WP, all these efficiencies are independent from its value (see Section 2.2) as their concerning aggregation models do not consider multiple bandwidth intervals.

Our results reveal the tradeoff that is between the efficiencies of aggregation and routing either when PLSM or STM is applied. Given the uniform distribution of edge and bandwidth level values, the aggregation efficiency shows a linear decrease in all cases. On the other hand, success ratio and delay efficiency show a considerable improvement as the number of bandwidth intervals increases especially for lower numbers, as the length of the bandwidth intervals decreases more significantly at smaller numbers of bandwidth levels.

Our results also show that STM outperforms the results that can be achievable by BP, WP and PLSM. It is also important to recognize that STM has a load balancing feature which helps keeping the blocking ratio low (thus the success ratio high). The reason for this property is the fact that the more heavily an aggregating edge is utilized, the higher average delay values its corresponding actual (physical) paths have. STM deals with this problem as it re-aggregates paths on the basis of measurement data, once the given aggregation threshold is exceeded. The extensive study of this property is a matter of further investigations. As a consequence, between scalability and routing efficiency, better compromises can be made by applying STM comparing to the cases when other methods are applied.

5. Conclusions

In our study, multi-domain routing has been discussed with focusing on the effect that topology aggregation has on different efficiency-related quantities. The qualities of scalability and routing efficiency have been considered through the evaluation of functions characterizing them. A Statistical Threshold-based Model (STM) has been introduced for topology aggregation which has outperformed the – otherwise highly efficient – Poly-line Segment aggregation Method (PLSM). Our simulation results show that there is indeed a tradeoff between scalability and routing efficiency, which can be be improved by applying our aggregation model.

There is a number of important advantages of STM:
- more accurate representation thus higher routing efficiency can be achieved,
- the computational time of the re-aggregation can be reduced significantly as the representatives do not need to be computed, and
- the method is flexibly tunable by setting the values of threshold parameters and the length of the FIFOs.

However, it also has a few drawbacks:
- additional storage requirements (FIFOs), and
- novel management tasks are necessary to implement.

Future directions of research should consider the optimal threshold values and size of data storage for such measurement-based statistical aggregation models. Examining the possible adaptation of this model within the framework of real networks and protocols is also of fundamental interest.

Author

ZOLTÁN CSERNÁTONY received his MSc degree in Electrical Engineering (with a final grade "excellent") in 2010 from the Budapest University of Technology and Economics (BME). His research interests include the scalability, efficiency and economical aspects of routing in multi-domain networks. Since his graduation, he has been admitted for PhD studies at the same institute to start working on his degree from September 2010 onwards.

References

[1] Mishra, "Scalability in communication networks", IEEE Network, 16(4), August 2002.

[2] T. Korkmaz and M. Krunz, "Source-oriented topology aggregation with multiple QoS parameters in hierarchical networks," ACM Trans. Modeling Comp. Simulation, Vol. 10, No. 4, pp.295–325, October 2000.

[3] R. Guérin and A. Orda, "QoS-based routing in networks with inaccurate information: theory and algorithms" IEEE/ACM Trans. on Networking, Vol. 7, pp.350–364, June 1999.

[4] B. Awerbuch and Y. Shavitt, "Topology aggregation for directed graphs," IEEE/ACM Trans. on Networking, Vol. 9, pp.82–90, February 2001.

[5] F. Hao and E. W. Zegura,
"On scalable QoS routing: performance evaluation of topology aggregation,"
In Proc. IEEE INFOCOM 2000, pp.147–156.

[6] Lui, K., Nahrstedt, K. and Chen, S.,
"Routing with topology aggregation in delay-bandwidth sensitive networks",
IEEE/ACM Trans. on Networking,
Vol. 12, Issue 1, pp.17–29.
February 2004.

[7] R. Hou et al.,
"Aggregation-Based QoS Routing in the Internet,"
Journal of Communications,
Vol. 5, No. 3, pp.239–246,
March 2010.

[8] Tang Y., Chen S., Ling Y.,
"State aggregation of large network domains",
Computer Communications,
Vol. 30, Issue 4, pp.873–885,
February 2007.

[9] L. Xiao, J. Wang, K.-S. Lui, K. Nahrstedt,
"Advertising Inter-domain QoS Routing Information,"
IEEE Journal on Selected Areas in Comm.,
22(10): 1949–1964,
December 2004.

[10] D. Jurca and R. Stadler,
"H-GAP: Estimating Histograms of Local Variables with Accuracy Objectives for Distributed Real-Time Monitoring",
IEEE Trans. on Network and Service Management,
Vol. 7, Issue 2, pp.83–95,
June 2010.

[11] Y. Yu, et al.,
"On the efficiency of inter-domain state advertising in multi-domain networks",
In Proc. of the 28th IEEE Conf. on Global Telecomm.,
2009.

[12] W. Lee,
"Topology aggregation for hierarchical routing in ATM networks,"
In ACM SIGCOMM Comp. Comm. Rev.,
Vol. 25, pp.82–92,
April 1995.

[13] M. Faloutsos, P. Faloutsos and C. Faloutsos,
"On power-law relationships of the internet topology,"
In Proc. ACM SIGCOMM 1999, pp.251–262.

[14] B. M. Waxman:
"Routing of multipoint connections,"
IEEE Journal on Selected Areas in Comm.,
Vol. 6, pp.1617–1622,
December 1988.

[15] J. Zhang, Y. Han, L. Wang,
"A New Topology Aggregation Algorithm in Hierarchical Networks",
ISECS Intl. Coll. on Comp., Comm., Ctrl. and Mgmt.,
2008, pp.179–183.