IP Traffic Engineering with Weighted Distributed Routing

Jonas Griem

University College London

Abstract: In this paper an introduction to QoS traffic engineering is presented. Particular emphasis is given to the scalability issues arising from state information that has to be stored in routers. It is found that while an MPLS based solution could potentially provide a more optimal results to QoS TE, it is quite unscalable as well as inflexible in case of failure. An IP based solution should be preferred if it can be made to behave similarly optimal to MPLS. Further work of this PhD will attempt to find appropriate algorithms to implement traffic engineering based routing decisions into IGPs like OSPF using the link weights as meta state.

1 Introduction

With the prospect of becoming the ubiquitous multi-service network of the future, the Internet needs to evolve to support services with guaranteed Quality of Service (QoS) characteristics. Some protocols and mechanisms have been defined, but these need to be tied closely in order to provide end to end QoS guarantees. This paper provides a review of some of the challenges that have been tackled and identifies some areas that require more attention.

2 Enabling QoS over IP

The Integrated Services (IntServ) rely on explicit request and reservation of resources for individual traffic flows. The Resource Reservation Protocol (RSVP) [1] defined in the frame of IntServ has the purpose of signalling resource requests, reserving capacity on the links along the path taken by a flow. IntServ does not scale well, as large amounts of state information have to be stored in the core network routers. It is mainly for this reason that IntServ is not usually considered for large networks, although it could provide the required quality guarantees.

The Differentiated Services framework (DiffServ) [2] was positioned as a more scalable solution for enabling QoS than the stateful IntServ. Each IP packet is classified with a DiffServ Code Point (DSCP), mapping traffic onto classes of the same quality treatment. Datagrams are policed and labelled at the edge of a DiffServ domain and then treated with DSCP specific Per Hop Behaviour (PHB) at every router. The Assured Forwarding (AF) and Expedited Forwarding (EF) PHBs can be used to provide some quantitative guarantees for per hop queue behaviour, such as per hop maximum delay. However, there is no end to end guarantee. While the policing and marking process at the domain border has to be carried out per flow, different flows with the same PHB inside the network are treated as aggregate. The solution is thought to be scalable, because of this aggregate treatment of flows in the core.

Although DiffServ protocols and mechanisms have been defined in order to achieve QoS guarantees, the framework does not provide an architecture for end to end QoS delivery. End to end quality also depends on the route that a flow is taking. DiffServ provides no means for route selection. The DiffServ mechanisms therefore have to be augmented through intelligent Traffic Engineering (TE) functionality. This functionality has to define the behaviour of hardware and software components in order to achieve QoS guarantees, while maintaining a near optimal utilisation of the networks resources. Even without multiple service classes today, maintaining a well balanced network utilisation is a problem.

3 QoS Traffic Engineering

The network can be dimensioned or engineered “off-line” according to the expected traffic matrix derived from projected and past demands. Unexpected traffic fluctuations can be dealt with through dynamic or “on-line” TE functions, which complement the “off-line” functionality. Techniques for “off-line” traffic engineering range from mass over-provisioning to non automated human intervention to completely automated resource provisioning cycles. Dynamic traffic engineering ranges from localised to centralised and is completely automated in order to allow fast reaction to change.

The success of traffic engineering algorithms depends greatly on the amount and accuracy of the supplied information on traffic flows and topology, as well as the options available to the algorithm. A distributed, router-embedded algorithm, can find local optimums in the network. An example for this is the constraint

shortest path first (CSPF) algorithm [3]. An "off-line" algorithm is executed more centralised at the management plane (this may yet be distributed on many servers across to the network), these "off-line" techniques are able to find more global optimums. An example for an "off-line" TE approach was developed as part of the IST-Tequila project [4].

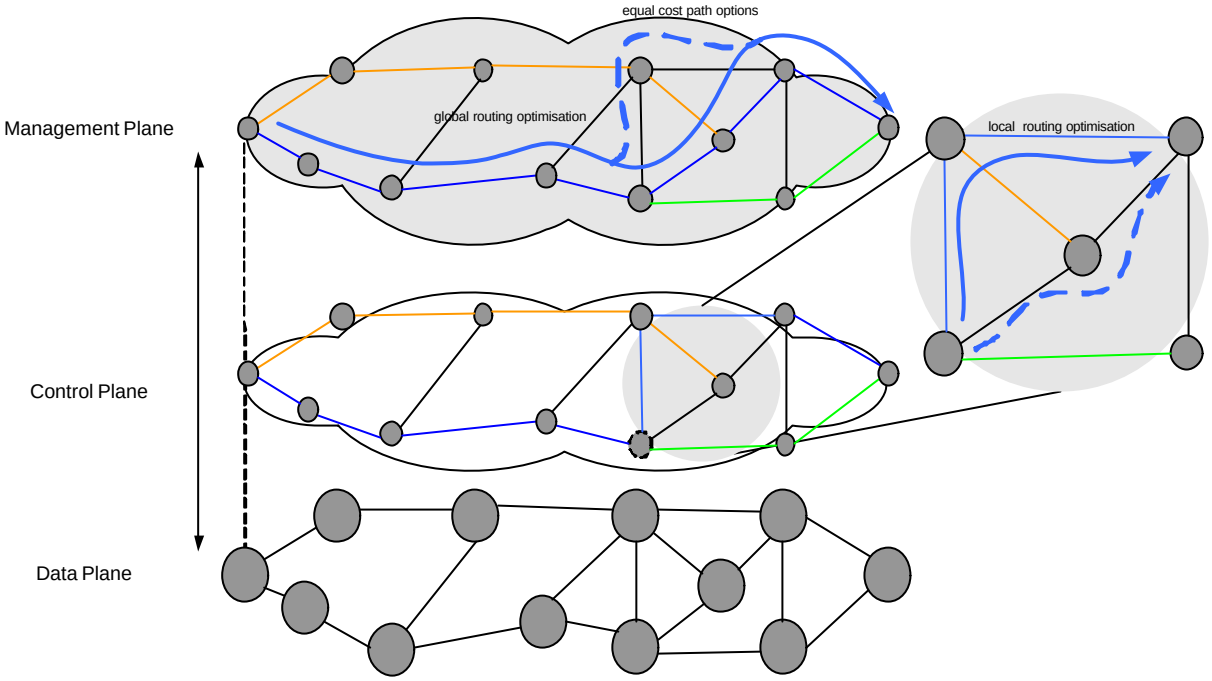

Figure 1 – Illustration of operation in management plane with global network view and control plane with local view. The global route selection and optimisation algorithm may leave options for a more frequently operating local algorithm to select equal cost path options that will reduce local congestion or underutilisation.

There are two options for enforcing routes that have been designed by the traffic engineering cycles in the network, stateful and stateless mechanisms. An example of each will be presented.

Multiprotocol Label Switching (MPLS) is a stateful mechanism, flows are tunnelled though the IP network on explicitly routed label switched paths (ERLSPs) [5]. MPLS is based on label translation, entries for each ERLSP have to be present in every router along the path. The label translation entries have to be set up before data transmission can begin. While MPLS itself provides bandwidth reservation, even with the DiffServ support defined in RFC3270 it does not provide other guarantees end to end. Most importantly however, MPLS provides the means for traffic engineering, as the physical path of an ERLSP is not limited to what the Interior Gateway Protocol (IGP) would choose as the shortest path to reach a destination. Traffic engineering based on MPLS has a number of consequences.

- The setup of ERLSPs allows control over the entire route, every hop can be explicitly defined. This means that the QoS management platform has complete control over the available resources, which is important for achieving hard guaranteed service.
- Each ERLSP requires state information to be present along the path. Consequently the path requires an initial set up, where appropriate label translation entries are made in routers. Traffic is then bound to this path explicitly, packet labels have no meaning outside the route. This is fundamentally different to routing with traditional IGPs.
- As all routes depend on label translation entries in routers, resilience to link failure is not naturally compensated as it is for the case of an IGP like Open Shortest Path First (OSPF). Explicit backup route configurations have to be made during initial set up, at the cost of storing more state information per ERLSP.

Stateless approaches like OSPF or Intermediate System to Intermediate System (IS-IS) routing do not suffer from these problems, but they come at a cost. The aggregate nature of traffic in these networks provides no direct means of individual per flow routing inside the domain. Routing is fundamentally limited to shortest path first calculations made by the IGP embedded in the routers and network

optimisation is thus limited. Traffic engineering can yet be done by setting the IGP link weights to influence the shortest path calculation of routes in such a way, that the routes follow a pattern as defined by the traffic engineering optimisation. Traffic engineering based on OSPF weight optimisation is not strictly stateless, as a certain amount of state is introduced in the routers.

4 Current Research

A number of papers have been published on optimisation of the link weights used for OSPF routing. These are usually set by simple algorithms, such as the inverse of the link capacity [6]. Using more sophisticated algorithms, [7] have shown that it is possible to get within 2-15% of optimality for network utilisation. The algorithms are based on local search heuristics, searching for local minima, while avoiding cycling and yet allowing for diversification so as not to converge on a sub-optimal local minimum. This research is particularly relevant for it provides important groundwork, such as cost metrics allowing to compare different style algorithms as well as proof that the idea is valid.

More recently, several papers have been published on the use of genetic heuristics in order to arrive at solutions still closer to the optimal solutions [8], than with the local search heuristics. Genetic algorithms model organic evolution based on the Darwinian principle of survival of the fittest. They are based on the principal three steps.

1. Randomly create an initial population P(0) of individuals.
2. Iteratively perform the following steps on the current generation of the population until the termination criterion has been satisfied.
    a. Assign fitness value to each individual using the fitness function.
    b. Select parents to mate.
    c. Create children from selected parents by crossover and mutation.
    d. Identify the best so far individual for this iteration of the algorithm.

A hybrid of genetic and other local search heuristic algorithms is thought to be a possibility for further optimisation of the weights problem.

5 Meta State Instead of per Flow State

The research presented provides a starting point for finding solutions leading to IP traffic engineering. So far it has been shown that OSPF can be optimised to utilisation close to the optimum achievable with MPLS [6]. However, complexity is added when introducing requirements for service guarantees. It is no longer sufficient to optimise the network to better utilise its capacity. QoS constraints of individual flows need to be taken into account. Hence a more finely grained control is required over individual flows, than has been achieved in research to date. The following gives a basic outline of an IP TE system to perform this task.

1. According to the expected traffic matrix derived from projected and past demands, a management based TE system performs “off-line” traffic engineering to arrive at an optimised network state.
2. The IP TE logic of the management plane derives suitable OSPF weights for routers on the network, translating the network state matrix resulting from the first step.
3. Weight settings are inserted into the routers at the end of the engineering cycle and OSPF is restarted.
4. Based on OSPF shortest path algorithms and on the weights set by the management system, the network should fall into the state that was calculated in the first step.
5. Dynamic traffic engineering algorithms embedded inside each router further optimise the weight settings, according to local optimisation strategies. These may also adopt to redistribute traffic in the case of congestion or underutilisation or redistribute multi path routes.

Compared to the MPLS based approach, instead of relying on per flow state information, this solution is based on aggregate “meta” states. The necessary state information is dramatically reduced and thus the approach is far more scalable. In addition, the meta state information is not necessary for network operation, it is responsible only for optimisation and QoS guarantees.

The IP based meta state system is hence robust in adaptation to link or node failure, which is handled by OSPF routing.

The proposal of dynamic traffic engineering embedded inside the routers could potentially be extended further. Whereas so far it was considered that the management solution accounts for each individual flow, the dynamic TE could potentially be self-optimising at a local level. It would require less precise information from the management platform, which would help improve scalability.

In order for a meta state IP TE approach become feasible, several problems need to be addressed.

- How much and what information does the weight calculation algorithm require in order to function adequately. This includes information about network topology, routing algorithms, etc.
- What is the information that has to be stored in the weight. It is likely that a more complex construct is required than is used in todays OSPF implementations.
- The stability of routes has to be ensured, all routers need to make consistent decisions, minor fluctuations in local weight settings may have a large impact on the overall network integrity (high sensitivity to minor changes in initial conditions).
- How computationally intensive is the algorithm and is full re-computation required for every newly set up route. The requirement for full re-computation may be a large setback, as large numbers of existing routes may have to be modified in order to accommodate new ones. A graceful restart of OSPF would be required in order to avoid upsetting existing QoS constraints.
- Along with dynamic TE comes the need to flood weight changes made locally to other routers. This potentially increases the amount of flooded information dramatically.
- Finally, how much less optimal is the approach compared to the MPLS approach. This consists of several measures, such as strictness of routes, network utilisation, etc.

6 Conclusion

After presenting requirements and a review of current research, the problem of extending weight optimisation research for OSFP into the area of QoS traffic engineering appears promising. None of the problems emerged thus far are simple and it may be sensible to start with simplified requirements. Building on the research done in the past, a first step could be to seek a more fine control of traffic patterns through weights rather than just optimising for best utilisation. If solutions can be found in this area, it is then possible to extend those to encompass multiple classes of DiffServ PHBs. Additionally the area of dynamic TE for local optimisation and congestion control seems a promising research topic.

7 Acknowledgements

The review presented in this paper and the initial work on IP traffic engineering have been carried out in the frame of the IST-MESCAL project. The author would especially like to thank David Griffin and Professor Chris Todd for their valuable input to the ideas presented here.

8 References

[1] J.Wroclawski, “The User of RSVP with IETF Integrated Services”, RFC 2210, IETF, Sep 1997

[2] S.Blake, D.Black, M.Carlson, E.Davis, Z.Wang, W.Weiss, “An architecture for Differentiated Services”, RFC 2475, IETF, Dec 1998

[3] “Junos Internet Software Configuration Guide: MPLS Applications”, http://www.juniper.net/techpubs/software/junos/junos57/swconfig57-mpls-apps/html/, Juniper, 2003

[4] P. Trimintzios, et al., Engineering the Multi-Service Internet: MPLS and IP-based Techniques, Proceedings of the IEEE International Conference on Telecommunications (ICT'2001), Bucharest, Romania, Vol. 3, pp. 129-134, IEEE, June 2001.

[5] E.Rosen, A.Viswanathan, R.Callon, “Multiprotocol Label Switching Architecture”, RFC 3031, IETF, Jan 2001

[6] D.Oran (Editor), “OSI IS-IS Intradomain Routing Protocol”, RFC 1142, IETF, Feb 1990

[7] B.Fortz, M. Thorup, “Internet Traffic Engineering by Optimising OSPF Weights”, 2000

[8] M.Ericsson, G.C.Resende, P.M.Pardalos, “A Genetic Algorithm For The Weight Setting Problem In OSPF Routing”, 2001