

TOWARDS AN OPTIMAL ROUTING STRATEGY

Vic Grout

*Faculty of Technology and Computer Science,
University of Wales,
NEWI Plas Coch, Wrexham, LL11 2AW, UK.
v.grout@newi.ac.uk*

ABSTRACT

The key features of the principal interior routing protocols for large systems are considered and compared, and the major weaknesses of the open standards noted. A proposal is given for an improved version (the *Enhanced Routing Algorithm - ERA*) and implementation options discussed. The results of initial simulation testing are summarized and future directions suggested in conclusion.

KEYWORDS

Link-state routing protocols

Partitioning

Optimization

1. INTRODUCTION AND DISCUSSION

An *Interior Routing Protocol (IRP)*, as opposed to an *Exterior Routing Protocol (ERP)* or *Border Gateway Protocol (BGP)*, is a Layer 3 protocol designed primarily to work within an *Autonomous System (AS)*; that is, a network or internetwork under a common administration. An important consideration is that an AS may be very large and optimality of routing across it, difficult to achieve. From the range of IRPs available, a few have features that make them particularly, or at least partially, suitable for use in large ASs. Three are considered initially here.

Open Shortest Path First (OSPF) is an open, i.e. non-proprietary, protocol in which the AS is divided (by the network administrator) into areas between which routing information, in the form of *Link-State Advertisements* and *Updates (LSAs and LSUs)* may be exchanged in summarised form making use of *Classless Inter-Domain Routing (CIDR)*, *Variable-Length Subnet Masking (VLSM)* and the partitioned nature of the AS. The term *link-state (LS)* implies that each participating station applying the protocol has a full knowledge of the topological state of the AS. Link-state routing protocols are generally more sophisticated than the alternative, *distance-vector (DV)* approach in which routers (say) are aware only of the direction and remoteness of target networks. We may use the term *node* as a general description of a routing station/router.

The *Enhanced Interior Gateway Routing Protocol (EIGRP)* is a Cisco Systems proprietary standard, which also works well in larger systems. Like OSPF, it permits CIDR/VLSM. However, in dispensing with the partitioned areas, EIGRP uses a *Diffusing Update Algorithm (DUAL)* to speed routing convergence. EIGRP is considered a hybrid (LS/DV) approach. A third protocol, *IS-IS (Intermediate System to Intermediate System)* has features in common with both OSPF and EIGRP and uses a hierarchy of partitioned areas.

Each option has its advantages and disadvantages. As a representative example, the strengths of OSPF may be summarised as:

- The principle of partitioning the AS into areas gives manageability of scale and permits reduced levels of routing traffic. This will continue to be the case when CIDR and VLSM have expired along with *Internet Protocol version 4 (IPv4)*, to be replaced with *Internet Protocol version 6 (IPv6)*.
- The route determination mechanism behind OSPF, *Dijkstra's Shortest Path Algorithm (SPA)* (Dijkstra 1959), is simple, easy to implement and polynomially computable.

However, it also has the following weaknesses:

- The areas, which give OSPF its ability to restrict the number and scale of routing updates, are assigned, not automatically, but by the network administrator. There is no mechanism for achieving or approximating partitioned optimality other than through manual control. In addition, these areas are essentially fixed from one configuration to the next.
- The manual implementation of OSPF is far from trivial. Even after the areas have been determined, the process of router configuration, particularly for *boundary* routers between areas, is time-consuming and has considerable scope for error. In fact, it requires some of the complexities of a BGP/ERP. Many network administrators even choose to forego the advantages of OSPF for such reasons.
- Dijkstra's SPA (DSPA) is optimal only on a pair-by-pair basis. It does not guarantee global optimality across the wider network, as the interaction among shortest paths for each node pair is not considered. As an example the standard cost function for OSPF, taken as the inverse of the link bandwidth, will in general, in conjunction with a pairwise SPA, generate combinations of routes sharing common links. The global optimum may be for individually longer routes to use independent paths across certain areas of the network to alleviate possible congestion.

A more complete treatment of these protocols is given in Aziz et al. (2002).

2. A NEW APPROACH

This paper proposes a new OSI Layer 3 protocol, the *Enhanced Routing Algorithm (ERA)*, having the following features:

- Automatic, and if appropriate dynamic, calculation of optimal or near-optimal areas.
- Simple implementation, eliminating the need to assign areas at configuration time.
- Globally optimised or optimally approximated routing.

Optimality, in the first and last points, is with respect to the efficiency of the dynamically derived routing strategy, which will come from taking a global rather than piecemeal approach to path determination. The first and second may be achieved through an automatic partitioning process. The third requires either a replacement for, or an alternative to DSPA.

2.1 Partitioning

There are a number of possible solutions to the problem of partitioning the nodes. One method is suggested in Grout (1988). We pursue an alternative approach here, based upon an older *minimal spanning tree (MST)* algorithm (Kruskal 1956). For this we need a measure of the *cost* of a link. In basic form, OSPF uses the inverse of link bandwidth as costs for DSPA. This is only one of a number of possibilities but serves well as an example.

ERA(1): Let there be n nodes. Let the data rate of the link between nodes i and j be b_{ij} . If i and j are not directly connected then $b_{ij} = 0$. Define the cost of the link (i, j) to be $c_{ij} = 1/b_{ij}$. If i and j are not directly connected then $c_{ij} = \infty$. Let d_{max} be the maximum *diameter* of a partition, the maximum distance between *any* two nodes in a partition. Initially define a set of partitions $\tilde{\mathbf{A}} = \{P_i\}$ where $P_i = \{i\}$ for each node i ; that is, each node is the sole member of its own partition. Set $d(P_i)$ (the diameter of the partition P_i) = 0 for each partition. Let c_{max} be the maximum link cost between two *adjacent* nodes in the same partition. Set the Boolean flag Pf ('partitions formed') to be false (0) at the outset. The optimal partitions may then be approximated as follows:

```

repeat
  if ( $\exists i', j' \in \mathcal{A} \text{ such that } c_{i'j'} = \min_{i,j \in \mathcal{A}} c_{ij}$  and  $c_{i'j'} \leq c_{max}$  and
       $P_{i'} \cap P_{j'} = \emptyset$  and  $d(P_{i'}, P_{j'}) \leq d_{max}$ ) then
    begin
       $P_{i'} := P_{i'} \cup P_{j'}$ ;
       $P_{j'} := \emptyset$ ;
    end
  else
    Pf := 1
until
  Pf

```

$\min_{i,j \in \mathcal{A}}$ is taken to mean the minimum over those values of i and j not already tried. \exists means 'there exists' and $\in \mathcal{A}$ 'such that'. At each iteration, the closest valid partitions are combined into a single, larger one; there are analogies with IS-IS. c_{max} limits the distance (cost) between adjacent nodes in the same partition and d_{max} the size of the partitions. The diameters $d(P_i)$ may be calculated either by counting links or adding costs. The algorithm is polynomial (in n) in complexity (Kershenbaum 1993).

2.2 Path determination

We now consider the calculation of routes. Suppose that partitions have been established and that nodes are exchanging LSAs. As an example we take the paths calculated by DSPA as a starting point although there are other possibilities (Ding-Zhu & Pardalos 1993 and Kershenbaum 1993).

ERA(2): Following the calculation of routes according to DSPA, define the boolean variable, x_{ab}^{ij} to be 1 if traffic between nodes i and j is carried by the link (a, b) and 0 otherwise. Then the *load* of the link (a, b) can be calculated as

$$l_{ab} = \sum_{i=1}^n \sum_{j=1}^n x_{ab}^{ij}$$

and its *weight* as $w_{ab} = l_{ab} c_{ab}$ (with c_{ab} defined as in the previous section). The weight of the complete network is then

$$W = \sum_{a=1}^n \sum_{b=1}^n w_{ab}.$$

Finally, let $W'_{(ij)}$ be the weight, recalculated as above, with the *second*-shortest path selected between i and j . Then the following will search for improvements to this initial solution:

```

repeat
  Mi := 0;
  for i := 1 to n do
    for j := 1 to n do
      if  $W - W'_{(ij)} > Mi$  then
        begin
           $i' := i$ ;
           $j' := j$ ;
           $Mi := W - W'_{(ij)}$ ;
        end;
      if  $Mi > 0$  then
        [recalculate]  $x_{i'j'}$ ;
until
  Mi = 0

```

(M_i – ‘maximum improvement’) Improvements will be found through re-routing traffic on heavily loaded links onto lighter loaded alternatives. This algorithm is also polynomial in n . This is a fairly crude approach, having features in common with EIGRP’s DUAL, and considers only the second shortest path for each node pair. It is presented in this form largely for the purposes of explanation. More sophisticated methods are available (Hershberger 2003).

3. RESULTS AND CONCLUSIONS

So, in principle, we have a compound algorithm (**ERA** = **ERA(1)** + **ERA(2)**) that delivers the requirements at the start of Section 2. There are many potential variations; only a single thread has been pursued here. In the variation given (**ERA(1)**), the application of a constrained form of Kruskal’s MST (**CKMST**) to the routing nodes delivers a set of optimal partitions that may be used as areas in an OSPF-like configuration. These areas could be static, simply removing the need for the network administrator to define them, or dynamic, calculated in response to changing network conditions. Also, as an example, it is shown (**ERA(2)**) how a sequence of perturbations/local searches may be used to improve upon an initial DSPA (say) solution by taking a global, rather than pair-by-pair view of routing optimality.

The results of initial experimentation and testing are encouraging. The partitioning process, applied with appropriate parameters, can be seen to deliver realistic groupings and the perturbation process will generally provide globally better routes than DSPA where such improvements exist. However, there are many variations and special cases to consider along with alternative methods of implementation. The necessary larger-scale and more extensive testing is continuing. Comments from other researchers are sought and welcomed.

A limitation of the process as presented here lies in the centralized nature of the component algorithms. A centralized routing strategy is generally accepted as being undesirable, it being preferable to allow route determination to take place on individual nodes. Work is proceeding to develop a fully distributed version of the **ERA** protocol in a form such as can be applied independently across network stations. A variation of CKMST (Prim 1957) is an option for **ERA(1)** and there are a number of potential solutions for **ERA(2)** (Träff 2000, for example). Both, however, lead to an interesting possibility: that each node’s view of network partitions might be different! With these *asymmetric partitions*, nodes send LSAs and route summaries according to their own perception of the logical network structure, but receive updates, etc. from other nodes in accordance with theirs. The concept is seen as intriguing and investigations continue.

REFERENCES

- Aziz, Z. et al, 2002. *IP Routing Protocols*. Cisco Press, USA.
- Dijkstra, E.W., 1959. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, Vol 1, pp269-271.
- Ding-Zhu, D. & Pardalos, P.M., 1993. *Network Optimization Problems: Algorithms, Applications and Complexity*. World Scientific, London.
- Grout, V., 1988. *Optimisation Techniques for Telecommunication Networks*, PhD Thesis, Plymouth Polytechnic, UK.
- Hershberger, J. et al, 2003. On the Difficulty of Some Shortest Path Problems, *Proceedings of the 20th Symposium on Theoretical Aspects of Computer Science*, Lecture Notes in Computer Science, Springer-Verlag, Berlin.
- Kershenbaum, A. 1993. *Telecommunications Network Design Algorithms*. McGraw-Hill, New York
- Kruskal, J.B., 1956. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. *Proceedings of the American Mathematical Society*. Vol. 7, pp48-50.
- Prim, R.C., 1957. Shortest Connection Networks and Some Generalizations. *Bell System Technical Journal*. Vol. 36, pp1389-1401
- Träff, J.L., 2000. A Simple Parallel Algorithm for the Single-Source Shortest Path Problem on Planar Digraphs, *Journal of Parallel and Distributed Computing*, Vol. 60, pp1103-1124.