RTGWG Working Group
INTERNET-DRAFT
Intended Status: Experimental RFC
Expires: September 2012

Shankar Raman
Balaji Venkat Venkataswami
Gaurav Raina
Vasan Srini

March 26, 2012

Power Based Topologies and Shortest Power Paths in OSPF
draft-mjsraman-rtgwg-ospf-power-topo-00

Abstract

In a Interior Gateway Protocol like OSPF (Open Shortest Path First) the computation of the shortest path tree to all destinations is computed for an area say a backbone or a non-backbone area. With importance given to the reduction of power within a network it becomes important to provide a solution that reduces the power consumed amongst routers and links that make up the network (in this case an area or a collection of areas including the backbone and non-backbone areas). This proposal aims at providing such a solution by producing a power topology of the area / areas. This power topology is constructed by assigning metrics to links based on the power consumed by the linecards (and hence their respective ports in an indirect way) of adjacent routers that are interconnected by each such link.

Status of this Memo

http://www.ietf.org/shadow.html

Table of Contents

1. Introduction

Estimates of power consumption for the Internet predict a 300% increase, as access speeds increase from 10 Mbps to 100 Mbps [3], [8]. Access speeds are likely to increase as new video, voice and gaming devices get added to the Internet. Various approaches have been proposed to reduce the power consumption of the Internet such as designing low-power routers and switches, and optimizing the network topology using traffic engineering methods [2].

1.1 Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

1.2 Low-power routers and switches

Low-power router and switch design aim at reducing the power consumed by hardware architectural components such as transmission link, lookup tables and memory. In [4] it is shown that the router's link power consumption can vary by 20 Watts between idle and traffic scenarios. Hence the authors suggest having more line cards and running them to capacity: operating the router at full throughput will lead to less power per bit, and hence larger packet lengths will consume lower power. The two important components in routers that have received attention for high power consumption are buffers and TCAMs. Buffers are built using dynamic RAM (DRAM) or static RAM (SRAM). SRAMs are limited in size and consume more power, but have low access times. Guido [1] states that a 40Gb/s line card would require more than 300 SRAM chips and consume 2:5kW. DRAM access times prevent them from being used on high speed line cards. Sometimes the buffering of packets in DRAM is done at the back end, while SRAM is used at the front end for fast data access. But these schemes cannot scale with increasing line speeds. Some variants of TCAMs have been proposed for increasing line speeds and for reduced power consumption [7].

1.3 Power reduction using routing and traffic engineering

At the Internet level, creating a topology that allows route adaptation, capacity scaling and power-aware service rate tuning, will reduce power consumption. In [8] the author has proposed a technique to traffic engineer the data packets in such a way that the link capacity between routers is optimized. Links which are not utilized are moved to the idle state. Power consumption can be reduced by trading off performance related measures like latency. For

example, power savings while switching from 1 Gbps to 100 Mbps is approximately 4 W and from 100 Mbps to 10 Mbps around 0:1 Watts. Hence instead of operating at 1 Gbps the link speed could be reduced to a lower bandwidth under certain conditions for reduced power consumption.

Multi layer traffic engineering based methods make use of parameters such as resource usage, bandwidth, throughput and QoS measures, for power reduction. In [6] an approach for reducing Intra-AS power consumption for optical networks that uses Djikstra's shortest path algorithm is proposed. The input to this method assumes the existence of a network topology using which an auxiliary graph is constructed. Power optimization is done on the auxiliary graph and traffic is routed through the low-power links. However, the algorithm expects the topology to be available for getting the auxiliary graph. While [6] handles optical networks and their corresponding power consumption, it does not take into account other link layer technologies. It is specialized for optical and not for heterogenous links that will exist in common OSPF domains.

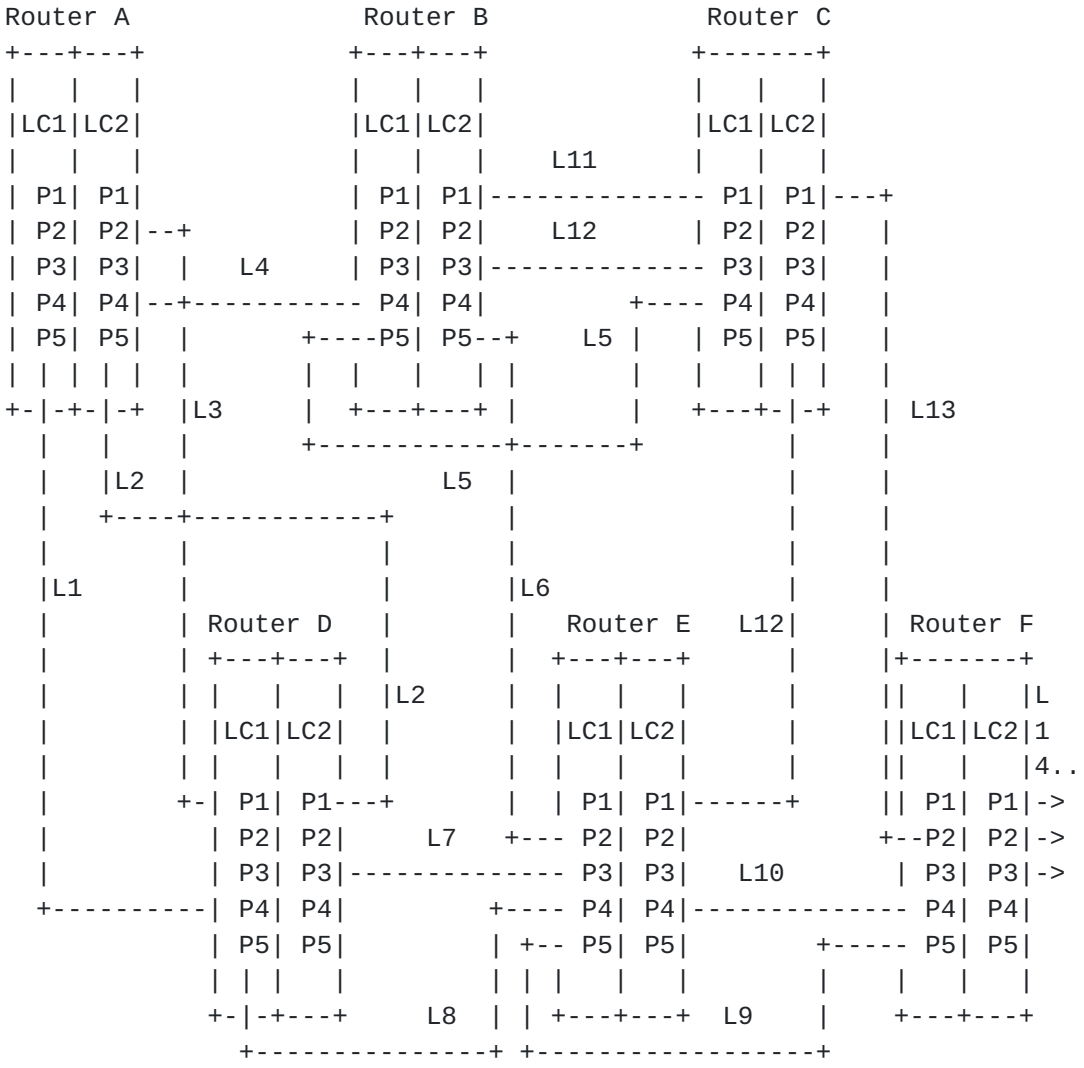
The proposal we make in this document indicates ways to solve the power reduction problem, by calculating a POWER metric whose importance is highlighted in the below mentioned sections. This POWER metric is obtained by including the factors such as power consumed by a linecard on a single chassis or multi-chassis router and consequently a port on that linecard by proportionally calculating power consumed for that port and hence for the link. The other factor that is taken into account is the utilization on that port and hence on that link.

2. Methodology

For each router / switch there exist linecards and each linecard has a set of ports or sometimes just one port of high capacity. This usually applies on routers and switches that are either single chassis or multi-chassis in their characterisation. By single chassis we mean that there exists a single chassis and slots for the Route Processor Card (one or more of these) typically upto to two of them, and one or more slots for linecards each having their respective characteristics such as number of ports (port density), type of such ports (SONET, ethernet, ATM etc..) usually depending on the link layer technology they support. Links are connections between ports on these linecards to other ports on linecards of other single chassis or multi-chassis system. A multi-chassis system is one that has multiple such chassis interconnceted amongst each other to form a single logical view of the system. Both single and multi-chassis have

linecards and respective ports on these linecards. Multi-chassis typically have a switch fabric chassis which connects each of these chassis to each other or to chassis of other multi-chassis or single chassis systems.

Consider the following topology...



The table of links between the various routers (which are assumed to be single chassis systems) is as follows...

Links	Routers	LC <> LC	Port Conn.	Capacity	Utiln. betw. 0..1
L1	A <> D	LC1<>LC1	P5<>P4	10G	.75
L2	A <> D	LC2<>LC2	P5<>P1	10G	.60
L3	A <> D	LC2<>LC1	P2<>P1	10G	.60
L4	A <> B	LC2<>LC1	P4<>P4	10G	.20
L5	B <> C	LC1<>LC1	P5<>P4	10G	.35
L6	B <> E	LC1<>LC1	P6<>P2	10G	.10
L7	D <> E	LC2<>LC1	P3<>P3	10G	.60
L8	D <> E	LC1<>LC1	P5<>P4	10G	.15
L9	E <> F	LC1<>LC2	P5<>P5	100G	.20
L10	E <> F	LC2<>LC1	P4<>P4	10G	.15
L11	B <> C	LC2<>LC1	P1<>P1	10G	.30
L12	E <> C	LC2<>LC2	P1<>P5	10G	.20
L13	C <> F	LC2<>LC1	P1<>P2	10G	.10
L14	F <> OA	LC2<>	P1<>		.20

In the above topology assume all point-to-point links between the routers. For now we will deal with P2P links alone and not venture into Broadcast Multi-access links or Non-Broadcast Multi-access links etc.. It is suffice to show how the scheme works for P2P links and then move more specifically to other types of networks to demonstrate this method of calculating the power topology of the network in the figure.

Each linecard consumes a certain amount of power and it is vendor dependent as to how the power consumed relates to the utilization on any of the links to which the linecard connects to. It is possible that the said topology of routers come from one vendor or from multiple vendors. It is assumed that the algorithm proposed will have the power consumed by a linecard available as a readable value in terms of W or kW or whichever measurable metric that is provided by the vendor.

It is possible that some of the Linecards are more capable than the others. Consider that Router A is a more capable router with more powerful linecard with higher port density. This is not shown in the figure, but assume so. LC1, LC2 on Router A could be consuming more power than the other Linecards on other routers. The main reason could be that LC1 and LC2 may have higher port density or higher

speed ports than the other routers. In order to calculate the power consumed on a link by a linecard it is important that we normalize the power as power consumed per port. Here the ports are normalized to lowest common denominator. If all links in the topology have 10G port capacity then the power calculated should be in terms power consumed per 10G port.

Assuming we have done this normalization we go on to calculate the POWER metric for each of the ports involved in a link which is derived as follows...

$$\text{POWER metric for a given Port on a LC} = \frac{\text{Power consumed per XG (normalized bandwidth) port}}{\text{Available Utilization on that port}}$$

Assume link L1. The ports concerned are both 10G and the ports are P5 on Router A and P4 on Router D. For calculating the POWER metric for a link which we will call PWRLINK we calculate the POWER metric for each side of the link and average the two to get PWRLINK.

$$\text{So PWRLINK for L1} = \frac{\text{POWER for P5 on LC1 on Router A} + \text{Power for P4 on LC1 on Router D}}{2}$$

The above can also be weighted if there is a multi-capacity port on one side of the link and not on the other. A multi-capacity link is one which provides multiple bandwidth capabilities such (1G/10G/100G) for example but auto-negotiates with other end to provide a lesser than highest capacity service.

The PWRLINK metrices once calculated are flooded in already defined OSPF-TE-LSA as an adapted TE-metric and is typically flooded as a link characteristic.

It is important to note that the denominator for POWER metric is Available Utilization on that port. The Available Utilization is measured in terms of intervals and not as discrete quantities. This is in order not to flood PWRLINK metrics into the OSPF area in LSAs very frequently as utilization may constantly change. The same applies to POWER metric as well.

Once the LSAs have been flooded the Routers run SPF on the tree with PWRLINKs assigned to the topology and calculate the PWRLINK based topology. This Power based topology can be used for forwarding high bandwidth streams and to optimally use power within the area.

The Utilization column shows the utilization of the link

corresponding to the row and column intersection a figure between 0 and 1. If utilized 100% then the figure shown will be 1 and if none then 0 and for the rest somewhere in the middle. This figure is used as the numerator in the POWER metric computation for that port.

2.1 Power Bias

Assume in the figure that there exist Routers A and D and that there is a bias on the link L1 in such a way that Router D computes a POWER metric of 10 and the Router D computes a POWER metric of 2 on the ports P5 and P4 respectively. Now the PWRLINK would be 6 for that link L1. Thus even if one side is excessively power guzzling then the PWRLINK moves up and thus is less preferred in the SPF tree and next hop computation for the Power topology.

If there is no bias and both the sides of the link are optimal in their power usage then the metric stays low even if more streams are sent on it. This is the main objective that is set out for router and switch manufacturers in the single chassis and multi-chassis world, in that they are incentivized to manufacture linecards that are not power hungry even if the number of packets flowing through them is high and thus the utilization is also high.

For those manufacturers who set a high power value for even minimal traffic, the vendors that dont would win out in the end.

2.2 ECMP links

It is possible that multiple links would have the same PWRLINK metric after a computation cycle. In such a case load-balancing techniques can be used to keep the ECMP links in a steady state with respect to each other. Depending on the utilization thereafter it is possible that the ECMP links may no longer be Equal cost but UCMP or Unequal Cost Paths.

2.3 Dampening the side effects of constant change

It is recommended in this draft that the implementation of the proposal be adaptive, infrequent in computation to the extent possible without sacrificing adapting to the dynamism and also reduce any frequent oscillations. The actual methods to adopt for this computation are outside the scope of this document.

3. Conclusion

Routers may have step levels in which they increase power consumption when they additively are loaded with more large bandwidth consuming multicast or unicast streams. Calibrating these levels may be useful

for implementing this scheme. It is possible that such calibrated thresholds can be used for advertising the PWRLINK ratios in the OSPF LSA advertisements. This would be useful for bringing down the frequency of updates or advertisements from a line-card about its PWRLINK ratio. When power consumption meanders within a certain given interval these ratios need not be re-advertised even if further unicast and/or multicast streams are added to it. The incentive is to recognize a linecard that does not drastically change power consumption even if large bandwidth streams are added onto it for forwarding and thus give it credit for its power optimal functioning. If a router tends to consume the highest level of power even when carrying low amounts of unicast and multicast streams on its line card, it would automatically have a poor ratio when compared to a router that efficiently uses power when considering the utilization being observed. The best case would be a low power consuming line-card or a router filled with such line cards that does not leave its power interval no matter how much ever capacity is sought to be used on it. But that would be an ideal condition but it is definitely an idealistic scenario towards which the router manufacturers should look at.

3 Security Considerations

<Security considerations text>

4 IANA Considerations

No new requirements are required from IANA for any new TLV as the TE-metric is adaptively changed to reflect the PWRLINK metric as well.

5 References

5.1 Normative References

[KEYWORDS] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC1776] Crocker, S., "The Address is the Message", RFC 1776, April 1 1995.

[TRUTHS] Callon, R., "The Twelve Networking Truths", RFC 1925, April 1 1996.

5.2 Informative References

[1] G. Appenzeller, Sizing router buffers, Doctoral Thesis, Department of Electrical Engineering, Stanford University, 2005.

[2] A. P. Bianzino, C. Chaudet, D. Rossi and J. L. Rougier, A survey of green networking research, IEEE Communications and Surveys Tutorials, preprint.

[3] J. Baliga, K. Hinton and R. S. Tucker, Energy consumption of the internet, Proc. of joint international conference on optical internet, June 2007, pp. 1-3.

[4] J. Chabarek, J. Sommers, P. Barford, C. Estan, D. Tsiang and S. Wright, Power awareness in network design and routing, Proc. of the IEEE INFOCOM 2008, April 2008, pp. 457-465.

[5] B. Venkat et.al, Constructing disjoint and partially disjoint InterAS TE-LSPs, USPTO Patent 7751318, Cisco Systems, 2010.

[6] M. Xia et. al., Greening the optical backbone network: A traffic engineering approach, IEEE ICC Proceedings, May 2010, pp. 1-5.

[7] W. Lu and S. Sahni, Low-power TCAMs for very large forwarding tables, IEEE/ACM Transactions on Computer Networks, June 2010, vol. 18, no. 3, pp. 948-959.

[8] B. Zhang, Routing Area Open Meeting, Proceedings of the IETF 81, Quebec, Canada, July 2011.

[9] M.J.S Raman, V.Balaji Venkat, G.Raina, Reducing Power consumption using the Border Gateway Protocol, IARIA conferences ENERGY 2012.

[10] A.Cianfrani et al., An OSPF enhancement for energy saving in IP Networks, IEEE INFOCOM 2011 Workshop on Green Communications and Networking

[EVILBIT] Bellovin, S., "The Security Flag in the IPv4 Header", RFC 3514, April 1 2003.

[RFC5513] Farrel, A., "IANA Considerations for Three Letter Acronyms", RFC 5513, April 1 2009.

[RFC5514] Vyncke, E., "IPv6 over Social Networks", RFC 5514, April 1 2009.

Authors' Addresses

Shankar Raman,
Department of Computer Science and Engineering,
I.I.T Madras,
Chennai - 600036
TamilNadu,
India.

EMail: mjsraman@cse.iitm.ac.in

Balaji Venkat Venkataswami,
Department of Electrical Engineering,
I.I.T Madras,

Chennai - 600036,
TamilNadu,
India.

EMail: balajivenkat299@gmail.com

Prof.Gaurav Raina
Department of Electrical Engineering,
I.I.T Madras,
Chennai - 600036,
TamilNadu,
India.

EMail: gaurav@ee.iitm.ac.in

Vasan Srini,
Department of Computer Science and Engineering,
I.I.T Madras,
Chennai - 600036
TamilNadu,
India.

EMail: vasan.vs@gmail.com