PANET Working Group
Internet-Draft
Intended Status: Experimental RFC
Expires: July 2013

Shankar Raman
Balaji Venkat Venkataswami
Gaurav Raina
Kamakoti Veezhinathan
IIT Madras

January 25, 2013

Reducing Power Consumption using BGP with power source data
draft-mjsraman-panet-inter-as-power-source-00

Abstract

In this paper, we propose a framework to reduce the aggregate power consumption of the Internet using a collaborative approach between Autonomous Systems (AS). We identify the low-power paths among the AS and then use Traffic Engineering (TE) techniques to route the packets along the paths. Such low-power paths can be identified by using the consumed-power-to-available-bandwidth (PWR) ratio as an additional constraint in the Constrained Shortest Path First (CSPF) algorithm. For re-routing the data traffic through these low-power paths, the Inter-AS Traffic Engineered Label Switched Path (TE-LSP) that spans multiple AS can be used. Extensions to the Border Gateway Protocol (BGP) can be used to disseminate the PWR ratio metric among the AS thereby creating a collaborative approach to reduce the power consumption. Since calculating the low-power paths can be computationally intensive, a graph-labeling heuristic is also proposed. This heuristic reduces the computational complexity but may provide a sub-optimal low-power path. The feasibility of our approaches is illustrated by applying our algorithm to a subset of the Internet. The techniques proposed in this paper for the Inter-AS power reduction require minimal modifications to the existing features of the Internet. The proposed techniques can be extended to other levels of Internet hierarchy, such as Intra-AS paths, through suitable modifications. The addition to this draft is that the power source of the Autonomous system is broken down to a ratio called PWR-SOURCE Ratio and used in the arrival of the metric to be used for this purpose.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering

Copyright and License Notice

Table of Contents

1 Introduction 4
1.1 Low-power routers and switches 4
1.2 Power reduction using routing and traffic engineering . . . 4
1.1 Terminology 5
2. Methodology 5
2.1 Pre-requisites for the Proposed Method 6
2.1.1 Constructing network topology using BGP strands 6
2.1.2 PWR ratio calculation 7
2.1.2.1 Power Sources as additional factor 9
2.1.2.2 Earlier method of computing numerator of PWR ratio. 10
2.1.3 Explicit routing using TE-LSPs 11
2.2 LOW-POWER PATHS 11

2.2.0.1 Algorithm 1 ASBR low-power path algorithm 12
2.2.0.2 Algorithm 2 PCE low-power path algorithm 12
2.2.1 Illustration 13
2.2.3 Equivalence class with total ordering 13
2.2.3.1 Algorithm 3 PCE low-power path algorithm with graph labeling 14
2.3 Implementation notes and Discussion 15
2.4 Applicability within ASes within a single Admin Domain . . . 18
2.4.1 PWR_SESSION 18
2.5 Conclusion and Future Work 19
2.5 Acknowledgements 22
3 Security Considerations 23
4 IANA Considerations 23
5 References 23
5.1 Normative References 23
5.2 Informative References 23
Authors' Addresses 24

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 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].

1.2 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. This topology is easy to obtain for Intra-AS scenario, but not for Inter-AS cases. In our approach, we propose a collaborative approach by AS in power reduction. The core of the Internet at the Inter-AS level, uses the Multi-Protocol Label Switching (MPLS) technology. MPLS label switched paths that traverse multiple AS carry traffic from a head-end to a tail-end. The AS use the Border Gateway Protocol (BGP) for exchanging routing and topology related information. One of the attributes of BGP namely, AS-PATH-INFO is used to derive the topology of the Internet at the AS level. The CSPF algorithm is run on this AS level topology with the consumed-power-to-available-bandwidth (PWR) ratio as a constraint, to determine the low-power path from the head-end to the tail-end. The PWR ratio can be exchanged among the collaborating AS using BGP. Explicit routing can be achieved between the head-end and the tail-end through the low-power paths connecting the AS using the Inter-AS Traffic Engineered Label Switched Path (TE-LSP) that span multiple AS.

Calculation of such low-power paths can be computationally intensive and hence certain heuristics may be needed to reduce the computation time. A graph-labeling heuristic is proposed to reduce the computation time, which may lead to sub-optimal low-power paths. We illustrate our approaches by applying it to a subset of the Internet topology. The rest of the paper is organized as follows: In Section II, we discuss in detail the pre-requisites for the algorithm. Section III introduces the proposed technique which uses the CSPF algorithm to calculate the low-power paths. We also show that by using a graph-labeling technique, we can reduce the computational complexity of the low-power path algorithm, but may obtain a sub-optimal low-power path. In Section IV, we discuss the implementation issues. We present our conclusion and future work in Section V.

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].

2. Methodology

<Document text>

2.1 Pre-requisites for the Proposed Method

In this section we discuss the pre-requisites for the implementation of the proposed scheme.

2.1.1 Constructing network topology using BGP strands

The Inter-AS topology can be modeled as a directed graph G = (V; E; f) where the vertices (V) are mapped to AS and the edges (E) map the link that connect the neighboring AS. The direction (f) on the edge, represents the data flow from the head-end to the tail-end AS. To obtain the Inter-AS topology, the approach proposed in [5] is used. In this approach, it is shown that a sub-graph of the Internet topology, can be obtained by collecting several prefix updates in BGP. This is illustrated in Figure 1 which shows the different graph strands of AS that are recorded from the BGP packets. Each vertex in this graph is assigned a weight according to the consumed-power-to-available-bandwidth (PWR) ratio of the AS, as seen by an Autonomous System Border Router (ASBR) that acts as an entry point to the AS. Figure 2 shows the strands merged together to form the topology sub-graph. In this figure, the weight of the vertices are mapped to the ingress edges. A reference AS level topology derived from 100 strands of AS-PATH-INFO received by an AS in the Internet is presented in Figure 3 in [9].

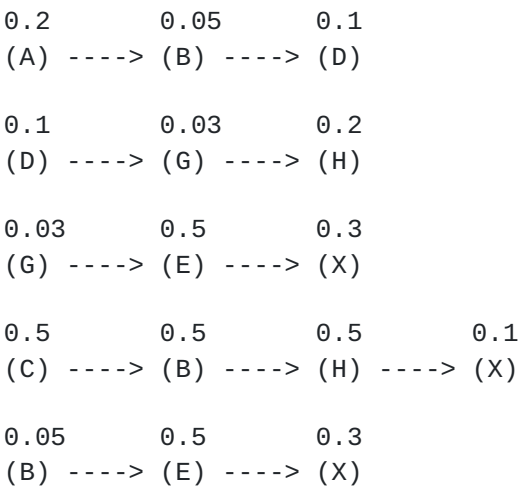

Figure 1: Different strands obtained from BGP updates, where vertices A,B,C,D and G represent the head-end AS. D,H and X form the tail-end AS. The vertex weights refer to the PWR ratio of the AS, and the direction of the link shows the next AS hop.

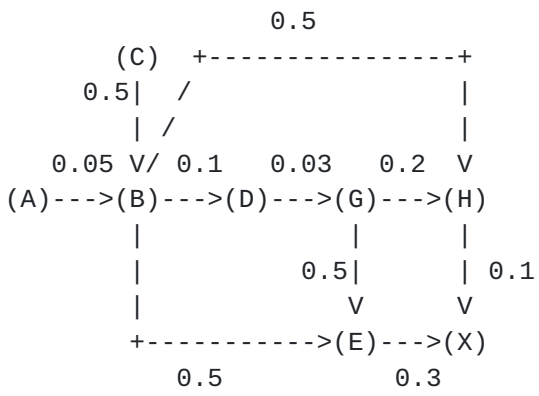

Figure 2:Combining the strands to get the topology of the Internet. The PWR ratio is mapped to the the ingress link of the ASBR and not to the AS.

2.1.2 PWR ratio calculation

In the topology sub-graph, each AS is expected to share its PWR ratio. In order to calculate this ratio we need to calculate the consumed power in the AS and the maximum bandwidth available with an ASBR.

In this proposal each AS is expected to share its PWR ratio from as many ASBRs (Autonomous System Border Routers) that it has. Intuitively in order to calculate this ratio we need to calculate the consumed power representative of the AS and the maximum bandwidth available with an ASBR on its egress links into the AS. The entry point to the AS is through the ASBRs that advertise the prefixes reachable through the AS. Hence the numerator of the PWR ratio is calculated for the AS at each ingress ASBR. We first obtain the summation of power consumed at the Provider (P) and the Provider Edge (PE) routers within an AS. The numerator of the PWR ratio is calculated by summing up the consumed power of all the routers to be taken into account and then dividing this sum by the number of routers. A more intuitive approach would be to use a weighted average method by assigning routers to categories and having appropriate co-efficients for each of these categories, thus arriving at a weighted average which is more accurate. One of these alternatives can be used to arrive at the numerator of the PWR ratio. Yet another alternative would have been to sum up the total consumed power of all routers in the AS and represent that as the numerator of the PWR ratio.

This average consumed power is divided by the maximum bandwidth available at each of the ASBR's egress link. This step is necessary

as the requested bandwidth for any path from the head-end to the tail-end using the ASBR is limited by the bandwidth available in the ASBR's egress links. The highest available bandwidth amongst the egress links of the ASBR is used as the denominator in the PWR ratio computation. If the entry point to the AS is through a different ASBR then the PWR ratio assigned to the ingress link of the ASBR might vary. Hence, an head-end AS might see different PWR ratios for an intermediate AS, if the intermediate AS has different ASBRs as its entry point.

The PWR ratio must be computed and disbursed much ahead of time before the Inter-AS TE-LSP explicit path or route is computed using the CSPF algorithm. The correctness of this ratio is of importance to compute the Inter-AS TE-LSP route through the green AS. If the entry point to the AS is through a different ASBR then the PWR ratio assigned to the ingress link of the ASBR might vary. Hence, an head-end AS might see different PWR ratios for an intermediate AS, if the intermediate AS has different ASBRs as its entry point.

We now illustrate the PWR ratio calculation. Consider an AS X which is one of the AS in the vicinity of another AS Y . Let this ASBR of X have 3 egress links into X denoted as E(1), E(2) and E(3), and 2 ingress links labeled I(1) and I(2). We now calculate the PWR ratio for I(1) and I(2). Assume that the routers in X have average consumed power of 200K Watts per hour. From figure 4 we can calculate the PWR ratio for I(1) and I(2) as 200K Watts / (60 * 60 * 1.5 Gigb = 3.7037 * (10 raised to -8) We could scale this to 0.37087 by multiplying with a base value of 10 raised to the 7th power.

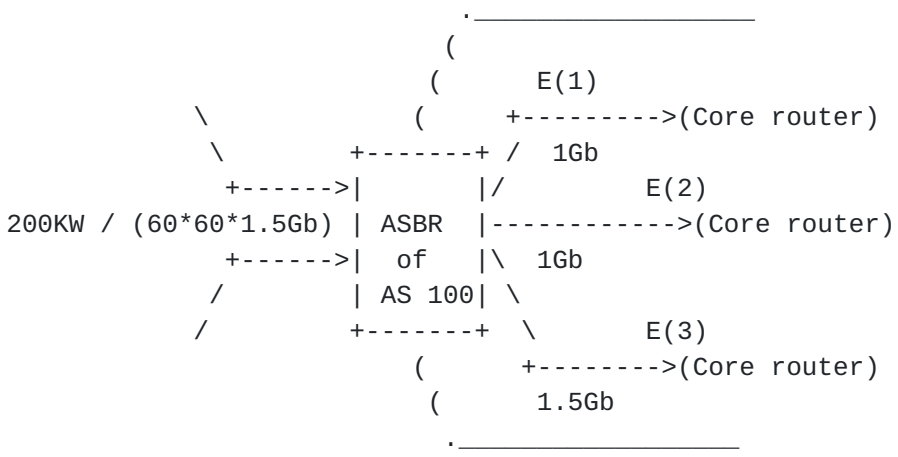

Figure 4:Calculation of PWR ratio by an ASBR associated with an AS. The I represents ingress links and E represents egress links. 200KW is the average consumed power in the AS. 1.5Gb is the maximum available bandwidth of the egress link in an ASBR.

Note that this ratio is actually a mapping function that is defined for each of the ingress links of the ASBR associated with an AS. For the head-end AS this mapping function does not exist as there is no ingress link. The PWR ratio can then be advertised to the other neighboring AS using the control plane through BGP extensions. BGP ensures that the information is percolated to other AS beyond the immediate neighbors. On receipt of these power metrics to the AS at the far-ends of the Internet, the overall AS level PWR ratio based Internet topology can be constructed. This view of the Internet is available with each of the routers without using any other complex discovery mechanism. Some sample link weights shown in Figure 2 is obtained by using such a mapping function on the ingress links.

2.1.2.1 Power Sources as additional factor

It is envisaged that the power sources of the Autonomous system using which the routers in the AS are powered should be declared as a metric which is further incorporated in the PWR ratio.

A suitable weight is provided to each type of source and the following table which is not claimed as totally exhaustive can be used to add this metric in the equation to compute the PWR ratio.

A formal classification of power sources and their weights is a topic to be considered later. For now we will deal with 2 main categories. Renewable sources of energy and non-renewable sources. There would be multiple categories under each of these major categories. Each such power source is assigned a weight.

Renewable Sources of Energy :

Wind - HighWeightOne
Solar - HighWeightTwo
Hydro - HighWeightThree
etc...

Non-renewable Sources of Energy :

Natural Gas - LowWeightOne
Petroleum and Diesel - LowWeightTwo
Nuclear - LowWeightThree
etc...

The PWR-SOURCE ratio is calculated in the proportion of how the above sources are combined to power the routers and its coolant systems and ancillary facilities in the AS.

Thus PWR-RATIO = (Consumed-Power / Available-Bandwidth)
* (1 / Weighted Average of Power Sources)

This compound metric could be used as the PWR metric in the calculations specified in this draft.

2.1.2.2 Earlier method of computing numerator of PWR ratio.

Earlier in the previous versions of this document in order to calculate this PWR ratio we needed to calculate the available power and the maximum bandwidth available with an ASBR. The entry point to the AS is through ASBRs that advertise the prefixes reachable through the AS. Hence, the numerator of the PWR ratio is calculated for the AS at each ingress ASBR. We first obtained the summation of power consumed at the major Provider (P) and Provider Edge (PE) routers within an AS. The average available power is obtained by subtracting the consumed power from the maximum power rating and summing the values for all the routers and then dividing the result by the number of routers. As an alternative, one could use a weighted average for more accuracy depending on the category of the router advertising the consumed power. Yet another alternative is to take the average or sum of the maximum power rating of all the routers within an AS without taking into account the consumed power. One of these alternatives was chosen to calculate the numerator of the PWR ratio.

Intuition however drives us towards consumed power as a better numerator since the lesser the power consumed the lesser the numerator and hence lesser the ratio if enough bandwidth is available at the ingress ASBR. The amount of consumed power per bit of

information ought to be low for the shortest path to work out properly. One more aspect is that lesser the power consumed per available bit of bandwidth it could be a sign that routers are more optimal in their power consumption as they take on more traffic. This is a very crucial point to be considered.

2.1.3 Explicit routing using TE-LSPs

We assume that the head-end and the tail-end may reside in different AS and the path is along multiple intervening AS. The way to generate this path is by using Traffic Engineered Label Switched Paths (TE-LSPs). TE-LSPs can influence the exact path (at the AS level) that the traffic will pass through. This path can then be realized by providing these set of low-power consuming AS to a protocol like Resource Reservation Protocol (RSVP). RSVP-TE then creates TE-LSPs or tunnels, using its label assigning procedure. The routers use these low-power paths created by the explicit routing method rather than using the conventional shortest path to the destination. By this way, we can influence exclusion of a number of high power AS on the way from the head-end to the tail-end AS. For example, the dotted line in Figure 5 represents the explicit route that is chosen by making use of such TE-LSPs from head-end AS A to the tail-end AS X. Note that if number of hop was the metric used by CSPF, then the route chosen is the path with 3 hops.

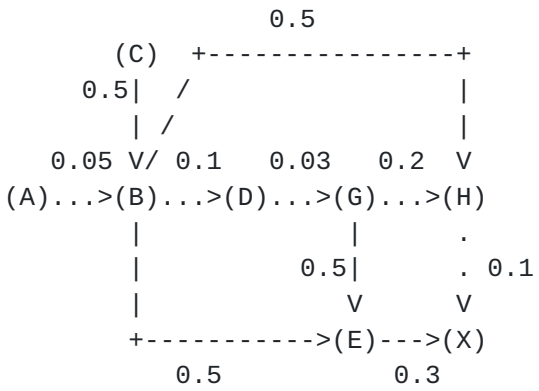

Figure 5:Low-power path is represented by the dotted lines. This low-power path has a longer number of hops than the conventional shortest path.

2.2 LOW-POWER PATHS

In this section we present the low-power path algorithm. The algorithm consists of two sub-algorithms: the first algorithm is

executed by all the ASBRs in the network and the second by all the Path Computation Elements (PCEs) in their respective AS. The algorithms for the ASBRs and PCEs are given as Algorithm 1, 2 and 3.

2.2.0.1 Algorithm 1 ASBR low-power path algorithm

Require: Weighted Topology Graph T=(AS, E, f)

1: Begin
2: if ROUTER == ASBR then
3: /* As part of IGP-TE */
4: Trigger exchange of available bandwidth on bandwidth change, to the AS internal neighbors;
5: BEGIN PARALLEL PROCESS 1
6: while PWR ratio changes do
7: Assign the PWR ratio to the Ingress links;
8: Exchange the PWR ratio with its external neighbors;
9: Exchange the PWR ratio with AS's (internal) ASBRs;
10: end while
11: END PARALLEL PROCESS 1
,br 12: BEGIN PARALLEL PROCESS 2
13: while RSVP packets arrive do
14: Send and Receive TE-LSP reservations in the explicit path;
15: Update routing table with labels for TE-LSP;
16: end while
17: END PARALLEL PROCESS 2
18: end if
19: End

2.2.0.2 Algorithm 2 PCE low-power path algorithm

Require: Weighted Topology Graph T=(AS, E, f)
Require: Source and Destination for Inter-AS TE LSP with sufficient bandwidth
1: Begin
2: if ROUTER == PCE then
3: Calculate the shortest paths from the head-end to the tail-end using CSPF with PWR ratio as the metric;
4: if no path available then
5: Signal error;
6: end if
7: if path exists then
8: Send explicit path to head-end to construct path;
9: end if
10: Continue passively listening to BGP updates to update T=(AS, E, f);
11: end if
12: End

2.2.1 Illustration

We now illustrate the proposed technique with a simple example. Consider the AS level topology sub-graph shown in Figure 5 constructed using the strands shown in Figure 1. The PWR ratio calculated at an ASBR which represents the metric for the AS is assigned to the ingress link. For example, AS H has two edges coming into it: one from B and the other from G. Note that the power metrics for the two strands are different as G to H is lower than that of B to H. This means that the lower power metric into H is better if the path from G to H is chosen rather than the one from B to H. This is illustrated in the Figure 5 using dotted lines. To construct a path with A as the head-end AS and X as the tail-end AS, from the AS level topology we see that the path A, B, H, X and A, B, E, X have the shortest number of hops. However by using CSPF with the PWR ratio metric as the constraint, we see that the path A, B, D, G, H, X is power efficient. The routing choice will however be based on the reservation of the bandwidth on this path. Given that available bandwidth exists to setup a TE-LSP, the explicit path A, B, D, G, H, X is chosen. The Resource Reservation Protocol (RSVP) adheres to its usual operation and tries to setup a path. If bandwidth is not available in the low-power path thus calculated, then we may fall back to other paths like A, B, H, X or A, B, E, X provided there is available bandwidth in these paths. The low-power path algorithm given as Algorithm 2 is executed by the PCE. Algorithm 1 prepares the topology and feeds it as input to the PCE as a weighted topology graph. Using the CSPF algorithm to calculate a route from a source to destination could be time consuming for a large networks. But the topology is dynamically updated and hence the computation of the shortest paths can be triggered based on need. We now give a heuristic method based on graph-labeling that reduces the computation time but could trade-off the optimal low-power path.

2.2.3 Equivalence class with total ordering

The heuristic is based on avoiding high PWR ratios. The approach partitions the weighted links into equivalence classes based on a range of PWR values. For each partition a labeling is applied such that each link in the partition has the same label. A total ordering relationship is then defined on the equivalence class. The heuristic then starts including partitions with minimum label value iteratively until we get a connected component, which includes the head-end and tail-end AS. We apply the CSPF algorithm with the weights as label values on this sub-graph to obtain the low-power path. The modified algorithm which uses this scheme is given in Algorithm 3. It should be noted that this algorithm could provide sub-optimal power paths as the intermediate steps carry incomplete Internet topology information.

2.2.3.1 Algorithm 3 PCE low-power path algorithm with graph labeling

Require: Weighted Topology Graph T=(AS, E, f)
Require: Source and Destination for Inter-AS TE LSP with
sufficient bandwidth
1: Begin
2: if ROUTER == PCE then
3: Group the links into N partitions with a label for
each partition depending on the PWR ratio
4: Sort the labels in ascending order.
5: repeat
6: Include the links that have the least label value;
7: Remove the partition with this label;
8: until there is a path from the head-end to tail-end AS
9: Calculate the low-power path using labels from the
head-end to the tail-end using CSPF ;
10: if no path available then
11: Signal error;
12: end if
13: if path exists then
14: Send explicit path to head-end to construct path;
15: end if
16: Continue passively listening to BGP updates to
update T=(AS, E);
17: end if
18: End

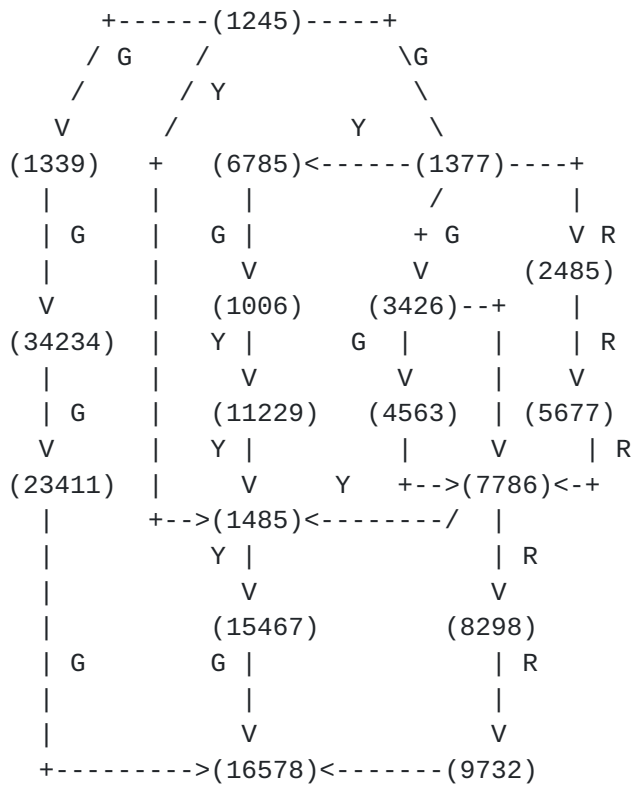

Figure 6:Application of the graph-labeling heuristic. We consider 3 labels "G" < "Y" < "R". Using algorithm 3 the "G" path from the head-end AS 1245 to the tail-end AS 16578 is chosen in the first iteration.

2.2.4 Illustration of graph labeling

We briefly illustrate the graph-labeling algorithm using Figure 6. In this diagram we have categorized the links into three partitions based on the PWR ratio. PWR ratio less than 0:1 are labeled as G, between 0:1 to 0:3 are labeled as Y and the rest as R. The total ordering is defined as G < Y < R, where the G links have low PWR ratios than the Y links. The path could be established through the AS that have G as the ingress link; the path being 1245, 1339, 34234, 23411 and 16578.

2.3 Implementation notes and Discussion

In this section we present some notes on feasibility of implementation of our scheme in a live network. First, the requested bandwidth should be available on the low-power path, but the CSPF algorithm is run with multiple constraints, one of which is the bandwidth requirement for the flows to be transported through the TE-LSP. The PWR ratio can then be applied to the available paths thus computing the low-power paths. Second, as we are using traffic

engineering with link state routing protocols, there is a reliable flooding process that are triggered when updates about the change in characteristic arise. We propose addition of some attributes with no change to the protocol implementation. There may be a time lag when the far ends of the Internet receive the attribute and the time it originated. This however cannot be avoided as with other attributes and metrics.

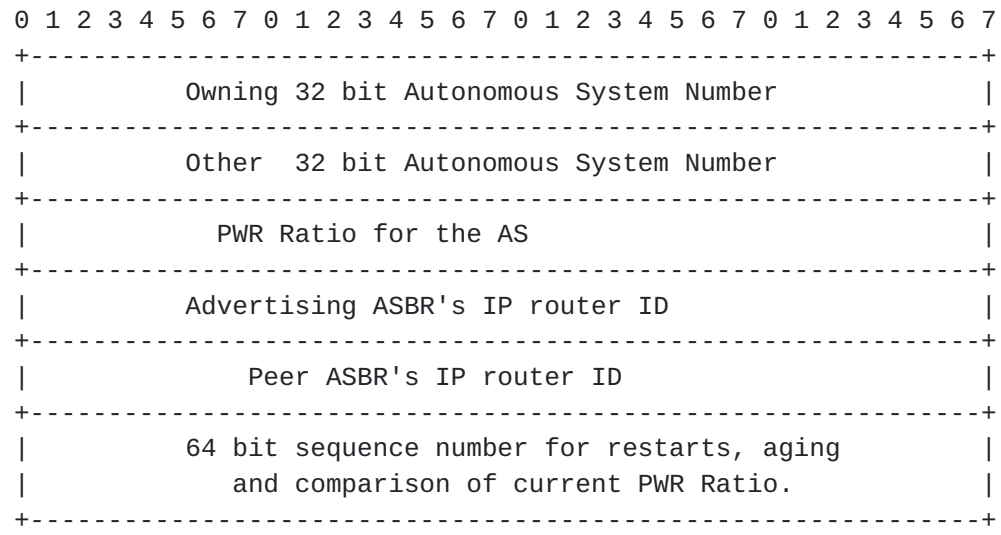

Figure 7: Proposed PDU format with an added attribute for AS-PATH-POWER-METRIC

The additions to the above Attribute have been added to optimize and correctly correlate the connecting ASes and the inter-AS links among them. For the traffic direction into the Advertising AS the above information will be easier to correlate than the previous version which did not advertise the peer AS which had the ingress links into the advertising Router.

In MPLS-TE when the TE metrics are modified, there is a reliable flooding process within an Interior Gateway Protocol (IGP). Such triggered updates apply to the PWR ratio as well. The proposed PWR ratio is advertised to the neighboring AS and the information percolated to all the AS, in a AS-PATH-POWER-METRIC attribute. This attribute can be implemented as shown in Figure 7. The frequency of the updates for this attribute should be fixed to avoid network flooding.

The AS-PATH-POWER-METRIC for each ASBR is calculated, and advertised as the PWR ratio for the AS. This AS-PATH-POWER-METRIC is filled into the appropriate optional transitive non-discretionary attribute and inserted into a unique vector for a set of prefixes advertised from the AS. Such advertised prefixes may have originated from the AS or

be the transit prefixes. The filled vector is sent to the ASBR of the neighboring AS and the information propagates to all the ASBRs. If the elements denoting AS in a vector of AS-PATH-INFO is not the same as the ones that need to be advertised in a AS-PATH-POWER-METRIC, then a suitable subset of AS-PATH-POWER-METRIC is identified and sent in the BGP updates. A vector of size 1 also can be employed if the AS in question is the only one for which PWR ratio has changed in the originating AS. The collation can be done depending on availability of such metrics and their mapping to a valid AS-PATH-INFO metric.

The power consumed by each router may fluctuate over short time intervals. In order to dampen these fluctuations which can cause unnecessary updates, power can be measured when falling within intervals of suitable size (say a range of values). This is as opposed to measuring power as a discrete quantity. This method of power measurement reduces the frequency of triggered updates from the routers due to power change.

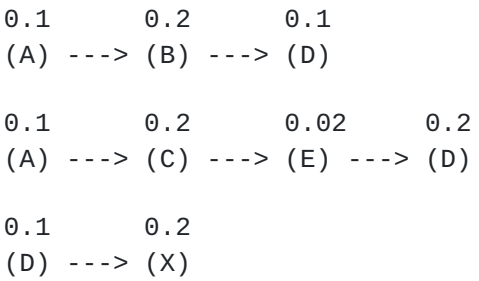

Figure 8: Example of strands where more than one PWR ratio is advertised by "D"

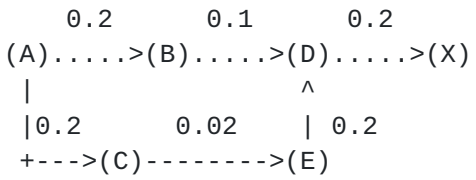

Figure 9:Choice of low-power path derived using the algorithm which uses lower value of the ingress link but through the same AS

A use case of multiple ASBRs advertising differing PWR ratio shows that an AS may be seen as green through one ingress link and not through the other. Consider the case of multiple ASBRs that belong to the same AS, advertising PWR ratios that differ. This could lead to power values that belong to different classes of ratios with many intervening classes in between. These advertised PWR ratios could lead to one ASBR being preferred over the other thus taking a different path from head-end to tail-end. This also entails that

there may be multiple paths to the AS through these different ASBRs.

Consider Figure 8 which shows a set of strands that derive a topology as in Figure 9. Here D is reachable via two paths but the PWR ratios differ. This illustrates the case where the better metric wins out. The average power consumed would not have an effect but the bandwidth available on these ASBR egress links would definitely influence the path.

2.4 Applicability within ASes within a single Admin Domain

As per [draft-ietf-idr-aigp] there are deployments in which a single administration runs a network which has been sub-divided into multiple, contiguous ASes, each running BGP. There are several reasons why a single administrative domain may be broken into several ASes (which, in this case, are not really "autonomous".) It may be that the existing IGPs do not scale well in the particular environment; it may be that a more generalized topology is desired than could be obtained by use of a single IGP domain; it may be that a more finely grained routing policy is desired than can be supported by an IGP. In such deployments, it can be useful to allow BGP to make its routing decisions based on the IGP metric, so that BGP chooses the "shortest" path between two nodes, even if the nodes are in two different ASes within that same administrative domain. The authors refer to the set of ASes in a common administrative domain as an "AIGP Administrative Domain".

A combination of the AIGP administrative metric and the graph heuristic algorithm could be combined to arrive at a set of a suitable number k power-shortest paths and then use a tie-break amongst such k power-shortest-paths with the least AIGP metric. This is provided the set of ASes where the decision is being made all fall under a AIGP Administrative domain. This provides a trade-off of power shortest paths and least number of hops (link wise) to get from source to destination across these ASes.

2.4.1 PWR_SESSION

An implementation that supports the PWR attribute CAN support a per-session configuration item, PWR_SESSION, that indicates whether the PWR attribute is enabled or disabled for use on that session.

- The default value of PWR_SESSION, for EBGP sessions, between providers (distinct operators) CAN be "disabled".

- The default value of PWR_SESSION, for IBGP and confederation-EBGP sessions, MUST be "enabled."

The PWR attribute MUST NOT be sent on any BGP session for which PWR_SESSION is disabled.

If an PWR attribute is received on a BGP session for which PWR_SESSION is disabled, the attribute MUST be treated exactly as if it were an unrecognized transitive attribute. That is, " The handling of an unrecognized optional attribute is determined by the setting of the Transitive bit in the attribute flags octet. Paths with unrecognized transitive optional attributes SHOULD be accepted. If a path with an unrecognized transitive optional attribute is accepted and passed to other BGP peers, then the unrecognized transitive optional attribute of that path MUST be passed, along with the path, to other BGP peers with the Partial bit in the Attribute Flags octet set to 1. If a path with a recognized, transitive optional attribute is accepted and passed along to other BGP peers and the Partial bit in the Attribute Flags octet is set to 1 by some previous AS, it MUST NOT be set back to 0 by the current AS".

This helps in confining the distribution of the attribute and use in calculation of the power shortest paths only amongst ASes that have trust relationships with other ASes. Of course, this includes and promotes the use of PWR attribute within a AIGP administrative domain.

2.5 Conclusion and Future Work

In this paper, we proposed a scheme for reducing the power consumption of the Internet using collaborative effort between AS. The topology of the Internet is represented using a graph model and derived using the strands obtained from the AS-PATH attribute of the BGP updates. CSPF algorithm is run on this topology by using the PWR ratio as a constraint. The PWR ratio is advertised through the ingress links of the ASBRs associated with AS using BGP updates. The CSPF algorithm finds out the low-power consuming AS that can route data packets from a head-end to a tail-end. Explicit routing is handled through the use of TE-LSPs. This entails adopting routes by choosing entry points to an AS that give energy saving paths. Since using CSPF can be time consuming a heuristic algorithm to derive the low-power paths using graph-labeling was proposed. Our work complements the current schemes for reducing power consumption within a router such as switching off or bringing to power-idle-state certain select components within the forwarding and lookup mechanisms.

This Power shortest Path calculation can be taken care of a Path Computation Element (PCE) unit that could be either be a process running on a linecard on a ASBR, or even a core router or an offline

engine that is passively listening to the BGP updates within the AS without spitting out any routes of its own. The PCE architecture has already been proposed in the ietf and even has a separate working group for itself.

These offline or linecard engines are currently being sold in the market by the networking majors and other companies that develop hardware and software for the PCE. All the PCE needs to do is to accept configuration and passively listen to BGP updates from various peers or even be a client for a route reflector, thus

a) Accepting these BGP updates

b) Extracting the AS PATH information from these updates

c) Then constructing the inter-AS topology

d) Apply the PWR metric that comes along in these BGP updates to the edges of the graph

e) Then compute the power shortest path as required by the configuration.

Normally the ASes have SLA agreements between each other to carry X amount of traffic from say a provider A. If the AS representing the ISP then advertises fake figures to carry more traffic than is mandated by the SLA agreement with other providers, then it is to that ISPs detriment since by advertising a better PWR ratio it invites more traffic through it thus getting paid less and carrying more traffic. This is not in the best interest of the ISP. This is so because in the final analysis the Power Shortest Path computed would include it regardless of the amount of traffic to be carried thus causing it to invite more traffic through it than it has accepted, even much more than its capacity. Hence it would be advisable for that ISP to advertise proper PWR ratios and NOT on the lower side of the spectrum. If it advertises HIGHER PWR ratios it would not be chosen, and hence that could be a policy measure NOT to accept any traffic at all since its capacity may be filled up with existing traffic. So advertising on the LOWER side would lead to lesser amount of benefit with respect to dollar per bit transported, and on the HIGHER side would be to exclude it from carrying any traffic that wanted to use the Power Shortest Path.

We also propose that there be a governing body in the IETF or outside it or sponsored by the IETF to verify the power ratios advertised are indeed valid or approximately closer to the actual consumption. A link up for each ISP with a power application level gateway to ensure

proper ratios are advertised could be mandated amongst at least the co-operating ISPs (ASes).

The points on which this proposal by us innovates is as follows.

a) There has been no effort prior to this to build an inter-AS topology with a weighted graph based on a PWR ratio. On this point it breaks a new path that would lead to inter-AS co-operation that contributes to power reduction overall in the internet. The paper suggested for OSPF by [10] deals with intra-Autonomous-system scenario rather than an inter-AS one. It is also to be noted that the IGP such as OSPF / IS-IS or any other link-state protocol for that matter is expected to capture the energy consumption of each router within the Autonomous system as in paper [10] to help get a hold on the overall average within the AS, or even sum up the total of all the power consumption within the AS with such intra-AS IGP LSA. This contributes to the PWR ratio proposed in our idea. Thus the intra-AS metric contributes to the PWR ratio. [10] proposal deals with primarily paths setup within an AS and not inter-AS paths. Thus the fundamental problem it solves is different while the problem we solve relates to the inter-AS paths which run across ASes from a head-end AS to a tail-end one.

b) The other aspect of innovation is to use BGP as the piggyback protocol upon which this scheme stands. There has been no effort earlier to approach the internet power reduction problem with BGP as the mode of transport of the energy ratios and coupling it with the inter-AS topology built with AS-PATH-INFO information.

The above 2 are key aspects of innovation.

When links and switches are gated or put into low-power state within an AS, the power-consumption automatically drops at the aggregate level, as a result of which the PWR ratio would be a lower figure advertised through BGP and thus this AS would attract more Power Shortest Path traffic through it. Thus the links within the AS and the switches within it would function more optimally if it had more traffic that went along paths that were originally put in low-power state thus utilizing the paths more effectively, when attracting PSP traffic.

There exist MIBs today that have object identifier for power consumed in a router. Maybe all the related components within it may NOT be listed with regards to power consumed. But the overall power consumed by the Router / Switch is gettable. Once it is advertised in a opaque Link-State-Advertisement say in the form of a TLV and the LSAs are flooded through the network in an AS, all routers get a uniform picture of which router consumes what power. This method already

exists for Traffic engineering Database LSAs that are advertised as LSAs for the purpose of traffic engineering within an AS. We are merely piggybacking on this capability to calculate the PWR ratio at the ASBR which amongst others is yet another Router / Switch of the AS.

Our future work includes looking into computing low-power paths within AS as well. Further it can be noted that the proposed algorithms might lead to increased latency as the number of hops increase, which could be critical for time sensitive applications. Since the PWR ratio could vary dynamically with traffic, the impact of traffic on the algorithm would also be of interest.

2.5 Acknowledgements

Shankar Raman would like to acknowledge the support by BT Public Limited (UK) under the BT IITM PhD Fellowship award. Balaji Venkat and Gaurav Raina would like to acknowledge the UK EPSRC Digital Economy Programme and the Government of India Department of Science and Technology (DST) for funding given to the IU-ATC. We would like to acknowledge that a version of this paper has been accepted in IARIA conference ENERGY 2012.

3 Security Considerations

No specific security considerations apart from the usual considerations with respect to authenticating BGP messages / updates from BGP neighbors is necessary for this scheme.

4 IANA Considerations

A new optional transitive non-discretionary attribute needs to be provided by IANA for carrying the PWR ratio across the Internet in the specified format in BGP.

5 References

5.1 Normative References

5.2 Informative References

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

[draft-ietf-idr-aigp] P. Mohapatra et.al, The Accumulated IGP metric attribute for BGP, https://datatracker.ietf.org/doc/draft-ietf-idr-aigp/, November 2012.

Authors' Addresses

Shankar Raman
Department of Computer Science and Engineering
IIT Madras
Chennai - 600036
TamilNadu
India.

EMail: mjsraman@cse.iitm.ac.in

Balaji Venkat Venkataswami
Department of Electrical Engineering
IIT Madras
Chennai - 600036
TamilNadu
India.

EMail: balajivenkat299@gmail.com

Prof.Gaurav Raina
Department of Electrical Engineering
IIT Madras
Chennai - 600036
TamilNadu
India.

EMail: gaurav@ee.iitm.ac.in

Prof.Kamakoti Veezhinathan
Department of Computer Science and Engineering
IIT Madras
Chennai - 600036
Tamilnadu
India

Email: kama@cse.iitm.ac.in