Reducing Power Consumption Using the Border Gateway Protocol

Shankar Raman*, Balaji Venkat‡ and Gaurav Raina‡
India-UK Advanced Technology Centre of Excellence in Next Generation Networks
**Department of Computer Science and Engineering, ‡ Department of Electrical Engineering*
Indian Institute of Technology Madras, Chennai 600 036, India
Email: mjsraman@cse.iitm.ac.in, balajivenkat@tenet.res.in, gaurav@ee.iitm.ac.in

***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 between the AS and then use Traffic Engineering techniques to route packets along the paths. Such low-power paths can be identified by using the available power-to-bandwidth (PWR) ratio as an additional constraint in the Constrained Shortest Path First algorithm. For re-routing the data traffic through these low-power paths, the inter-AS Traffic Engineered Label Switched Path that spans multiple AS can be used. Extensions to the Border Gateway Protocol 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-labelling 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.**

***Keywords*-Border Gateway Protocol; Autonomous systems; Traffic engineering.**

I. Introduction

Estimates of power consumption for the Internet predict a 300% increase, as access speeds move from 10 Mbps to 100 Mbps [2]. 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 approaches [3].

Low-power router and switch design aim at reducing the power consumed by hardware components such as transmission link, lookup tables and memory. In [7], it is shown that the link power consumption can vary by 20 Watts between idle and traffic scenarios. Hence, the authors suggest having more line cards and fully utilize them. Operating at full throughput will lead to less power per bit. Therefore, larger packet lengths will consume lower power. The two important components that have received attention for high power consumption are static and dynamic RAM-based buffers (SRAM, DRAM) and Ternary Content Addressable Memories (TCAM). A 40 Gb/s line card would require more than 300 SRAM chips and consume 2.5 kW [1]. Some variants of TCAMs have been proposed for high speed lines with reduced power consumption [10]. But these schemes cannot scale forever.

At the Internet level, creating a topology that allows route adaptation, capacity scaling and power-aware service rate tuning, will reduce power consumption. In [9], a subset of IP router interfaces are put to sleep, using an Energy Aware Routing (EAR) after calculating shortest path trees of the network from each router. Such a technique is useful in setting up paths within an Autonomous System (AS). In [5], the authors provide a way to introduce hardware standby primitives and apply traffic engineering methods to coordinate and reduce power consumption under given network operational constraints. Power savings while switching from 1 Gbps to 100 Mbps is approximately 4 W and from 100 Mbps to 10 Mbps around 0.1 W. Hence, instead of operating at 1 Gbps the link speed could be reduced to a lower bandwidth under certain conditions for reduced power consumption. A detailed review on energy efficiency of the Internet is given in [6].

Multilayer traffic engineering based methods make use of parameters such as resource usage, bandwidth, throughput and Quality of Service (QoS) measures, for power reduction. In [13], an approach for reducing intra-AS power consumption for optical networks using Djikstra's shortest path algorithm is proposed. The input assumes the existence of a network topology for constructing an auxiliary graph. This topology is easy to obtain for intra-AS scenario. Traffic is then re-routed through the low-power optimized links.

We propose a collaborative approach that uses inter-AS power reduction. Multi-Protocol Label Switching (MPLS) label switched paths that traverse multiple AS carry traffic from a head-end to a tail-end. 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 Constrained Shortest Path First algorithm (CSPF) uses the AS level topology with available power-to-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 and tail-ends through the low-power paths connecting the AS using inter-AS Traffic Engineered Label Switched

 ISBN: 978-1-61208-189-2

Path (TE-LSP) that span multiple AS. Since calculation of such low-power paths can be computationally intensive certain heuristics may be needed to reduce the computation time. A graph-labelling 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 uniqueness of our approach is that it can be used for inter-AS power reduction and requires cooperative effort from Internet Service Providers (ISP). Further, we use BGP the existing protocol in the Internet not only for detecting the topology but also to exchange power information and then construct low-power path.

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 for calculating the low-power path. We also show that by using a graph-labelling 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.

II. Pre-requisites for the proposed method

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

A. Constructing network topology using BGP strands

The inter-AS topology can be modelled 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 neighbouring 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 [12] 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 an AS recorded from the BGP packets. Each vertex in this graph is assigned a weight according to the available power-to-bandwidth (PWR) ratio of the AS, as seen by an Autonomous System Border Router (ASBR) that acts as an entry point. Figure 2 shows the merged strands forming the topology sub-graph where 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. For a detailed discussion on completeness of Internet topology information using BGP refer to [8], [11]. Any other algorithm that gives a complete AS topology could also be used.

B. PWR ratio calculation

In the topology sub-graph, each AS shares its PWR ratio. To calculate this ratio we need the available power and maximum bandwidth with an ASBR.

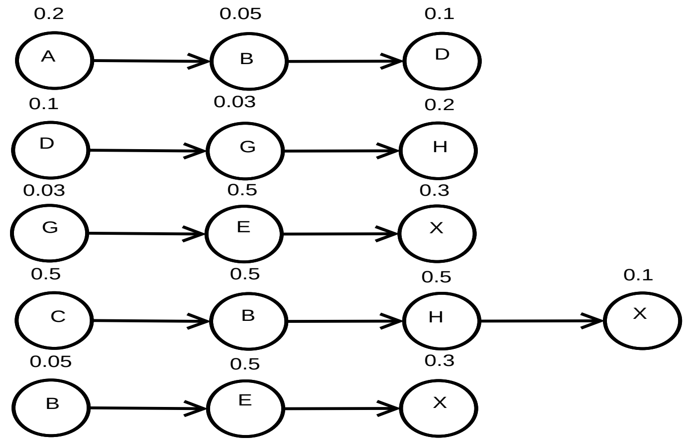

Figure 1. Strands obtained from BGP updates, vertices A,B,C,D and G are the head-end AS; D,H and X are the tail-end AS. The vertex weights represent the PWR ratio of an AS, and the link direction shows the next AS hop.

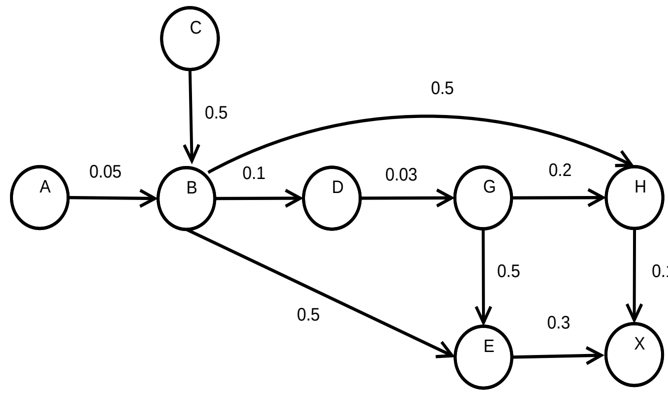

Figure 2. Strands combined to get the Internet topology. The PWR ratio is mapped to the ingress link of the ASBR.

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 obtain the summation of power consumed at the major Provider (P) and Provider Edge (PE) routers within an AS. These can be obtained by using any of the intra-AS power calculation techniques. The average available power is obtained by subtracting the consumed power from the maximum power rating, summing the values for all the routers and then dividing the result by the number of routers. Other alternatives include using a weighted average depending on the category of the router advertising the consumed power, or to take the average or sum of the maximum power rating of all the routers within an AS. The average available 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

 ISBN: 978-1-61208-189-2

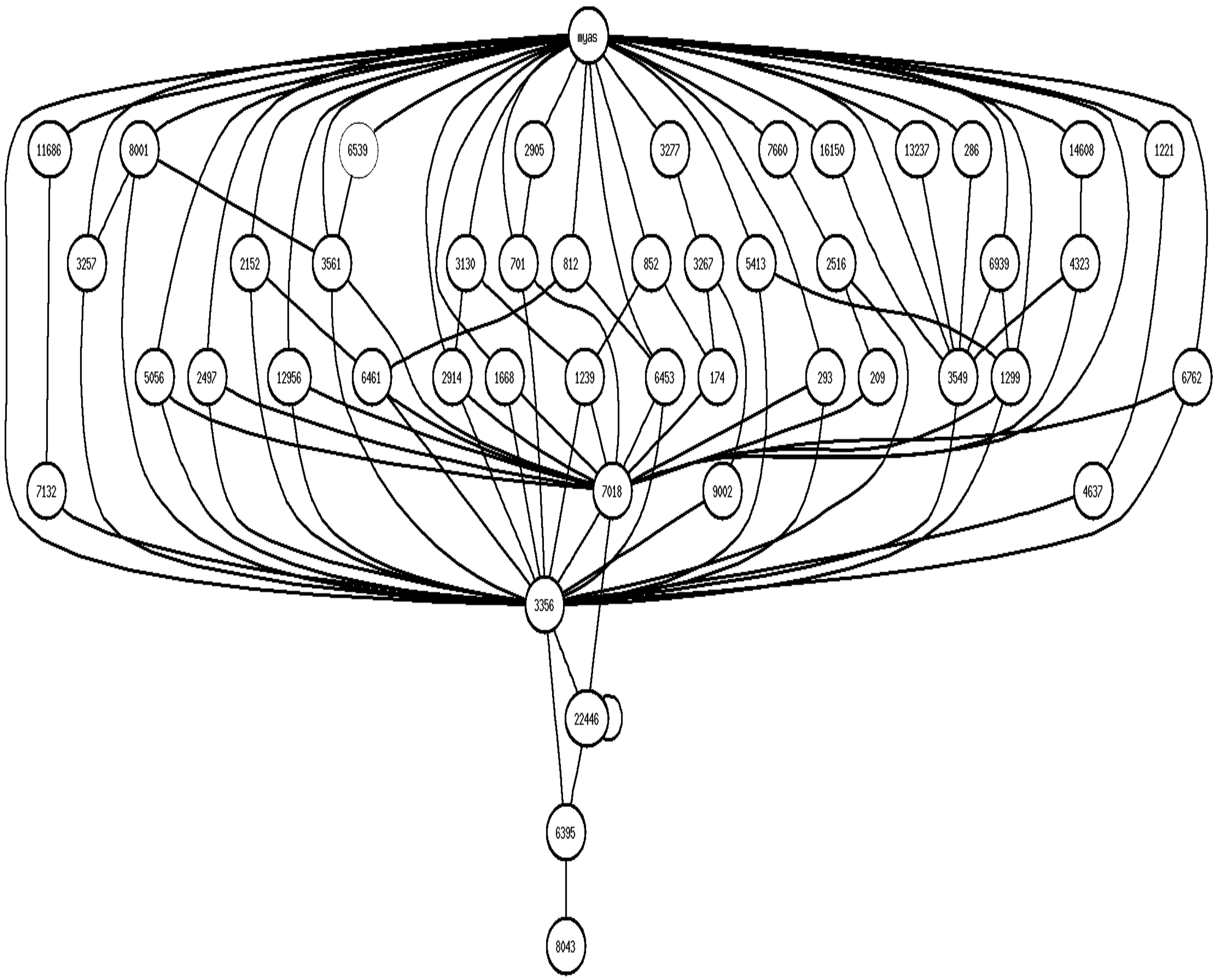

Figure 3. Internet topology graph derived from 100 strands of AS-PATH-INFO attribute by an AS through an ASBR. The top-most node (myas) represents the head-end and the bottom-most node (8043) represents the tail-end AS.

links. Simple Network Management Protocol can be used to extract this power information [4].

The highest available bandwidth amongst the egress links of the ASBR is used as the denominator in the PWR ratio computation. This PWR ratio must be computed and disbursed much ahead of time before the inter-AS TE-LSP explicit path is computed using the CSPF algorithm. The correctness of this ratio is of importance to compute the inter-AS TE-LSP route through the low-power 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, it is possible that an head-end AS might see different PWR ratios for an intermediate AS.

As an illustration, 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 denoted as $E(1)$, $E(2)$ and $E(3)$, and 2 ingress links labelled $I(1)$ and $I(2)$. We now calculate the PWR ratio for $I(1)$ and $I(2)$. Assume that the routers in X have average available power of 200 kW/hour. From Figure 4 we can calculate the PWR ratio for $I(1)$ and $I(2)$ as 200 kW/$(60 * 60 * 1.5$ Gb$) = 3.7037 * 10^{-8}$. We could scale this to 0.37037. This ratio is a mapping function defined for each of the ingress link of the ASBR of an AS. Note the absence of ingress link for the head-end AS.

The PWR ratio can then be advertised to the other neighbouring AS through the control plane using BGP extensions. BGP ensures that the information is percolated to other AS. On receipt of this PWR ratios by the AS at the far-end of the Internet, the overall AS level topology can be constructed. Note that view of the Internet is available with each of the routers without using any other complex discovery mechanism. Some sample link weights shown in

 ISBN: 978-1-61208-189-2

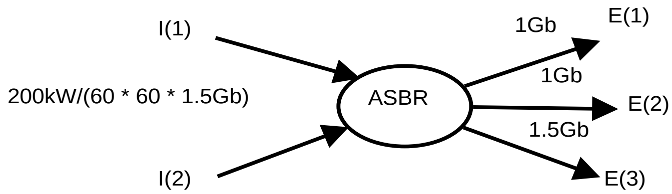

Figure 4. Calculation of PWR ratio by an ASBR of an AS. The I's are ingress links and E's are egress links. 200 kW/hour is the average available power in the AS. 1.5 Gb is the maximum available ASBR egress link bandwidth.

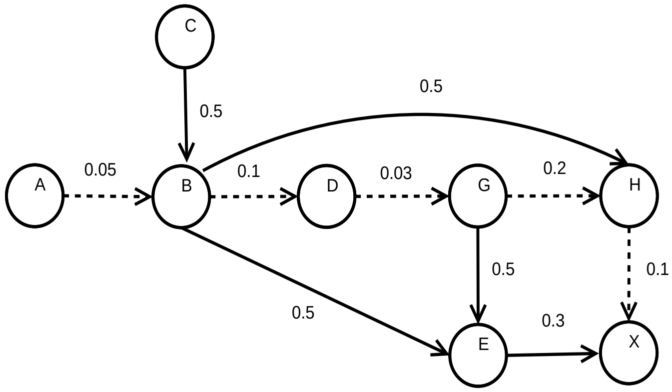

Figure 5. Dotted lines represent low-power path but has a longer number of hops than the shortest path.

Figure 2 are obtained by using such a mapping function on the ingress links.

C. Explicit routing using TE-LSPs

The head and tail-ends may reside in different AS and the path could span multiple intervening AS. To generate this path we can use Traffic Engineered Label Switched Paths (TE-LSPs). TE-LSPs can influence the exact path (at the AS level) for the traffic and this path can be realized by providing a 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 algorithm. This influences the exclusion of a number of high power AS on the path 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 the metric used is the number of hops, then the route chosen could be different.

III. LOW-POWER PATHS

In this section, we present the low-power path calculation algorithm. The algorithm consists of two sub-algorithms: the first algorithm is executed by all the ASBRs in the network and the other by all the Path Computation Elements (PCEs) in their respective AS. PCEs have been proposed by the Internet Engineering Task Force (IETF) for path computation activities. We can use the existing PCE architecture for our algorithm. The algorithms for the ASBRs and PCEs are given as Algorithm 1, 2 and 3.

Algorithm 1 ASBR low-power path algorithm

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

Begin
if ROUTER == ASBR **then**
  /* As part of IGP-TE */
  Trigger exchange of available bandwidth on bandwidth change, to the AS internal neighbours;
  BEGIN PARALLEL PROCESS 1
  while PWR ratio changes **do**
    Assign the PWR ratio to the Ingress links;
    Exchange the PWR ratio with its external neighbours;
    Exchange the PWR ratio with AS's (internal) ASBRs;
  end while
  END PARALLEL PROCESS 1
  BEGIN PARALLEL PROCESS 2
  while RSVP packets arrive **do**
    Send and Receive TE-LSP reservations in the explicit path;
    Update routing table with labels for TE-LSP;
  end while
  END PARALLEL PROCESS 2
end if
End

A. Illustration

We illustrate the 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 is assigned to the ingress link. 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. "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 "B" to "H". The dotted lines in Figure 5 represent low-power path.

To construct a path with "A" as the head-end and "X" as the tail-end in the AS level topology the paths "A", "B", "H", "X" and "A", "B", "E", "X" have the same number of hops. However by using CSPF with the PWR ratio as the constraint, the path "A", "B", "D", "G", "H", "X" is power efficient. The routing choice will depend on the reservation of the bandwidth on this path. If available bandwidth exists to setup a TE-LSP, then the explicit path

 ISBN: 978-1-61208-189-2

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

```
Begin
if ROUTER == PCE then
    Calculate the shortest paths from the head-end to the
    tail-end using CSPF with PWR ratio as the metric;
    if no path available then
        Signal error;
    end if
    if path exists then
        Send explicit path to head-end to construct path;
    end if
    Continue passively listening to BGP updates to update
    T=(AS, E, f);
end if
End
```

"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 shortest paths, provided there is available bandwidth. 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 the route from source to destination could be time consuming for large networks. But the topology is dynamically updated and hence the computation of the shortest path can be triggered based on need. We now give a heuristic method based on graph-labelling that reduces the computation time but could trade-off the low-power path.

B. Equivalence class with total ordering

The heuristic is based on avoiding high PWR ratios by partitioning the weighted links into equivalence classes based on a range of PWR values. For each partition a label 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 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 as Algorithm 3. It should be noted that this algorithm could provide sub-optimal power paths as the intermediate steps carry incomplete Internet topology information.

Algorithm 3 PCE low-power path algorithm with graph labelling

Require: Weighted Topology Graph T=(AS, E, f)
Require: Source and Destination for inter-AS TE LSP with sufficient bandwidth

```
Begin
if ROUTER == PCE then
    Group the links into "N" partitions with a label for
    each partition depending on the PWR ratio
    Sort the labels in ascending order.
    repeat
        Include the links that have the least label value;
        Remove the partition with this label;
    until there is a path from the head-end to tail-end AS
    Calculate the low-power path using labels from the
    head-end to the tail-end using CSPF ;
    if no path available then
        Signal error;
    end if
    if path exists then
        Send explicit path to head-end to construct path;
    end if
    Continue passively listening to BGP updates to update
    T=(AS, E);
end if
End
```

C. Illustration of graph labelling

We briefly illustrate the graph-labelling algorithm in Figure 6. In this figure, the links are categorized into three partitions based on the PWR ratio. PWR ratio less than 0.1 are labelled as "G", between 0.1 to 0.3 are labelled as "Y" and the rest "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 has "G" as the ingress link; the path being 1245, 1339, 34234, 23411 and 16578.

IV. 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 gets 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

 ISBN: 978-1-61208-189-2

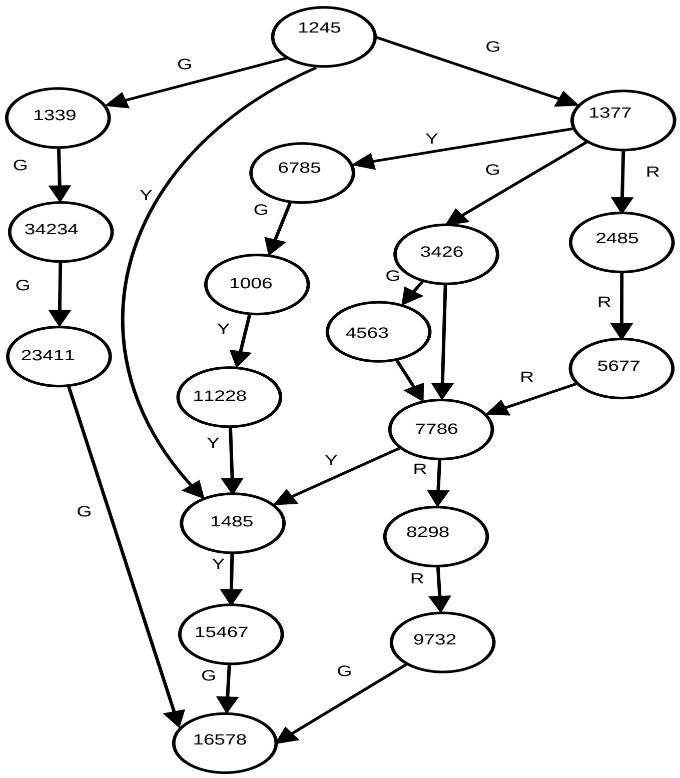

Figure 6. Application of the graph-labelling heuristic. We consider 3 labels "G" < "Y" < "R". Using algorithm 3 the "G" path from the head-end 1245 to the tail-end AS 16578 is chosen in the first iteration.

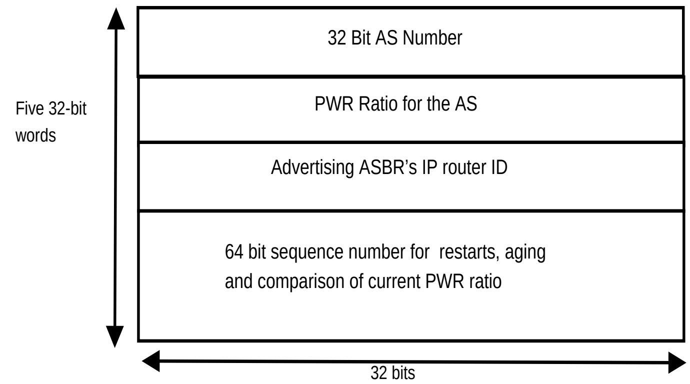

Figure 7. Proposed PDU format with a new attribute for AS-PATH-POWER-METRIC.

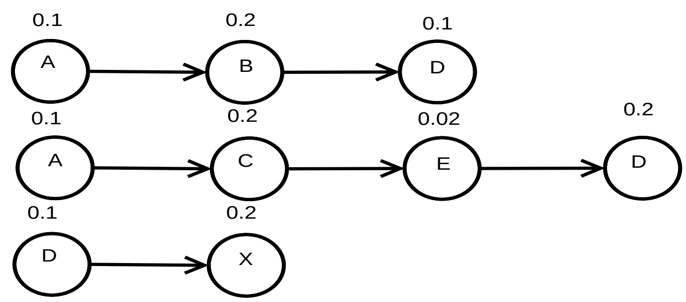

Figure 8. Example of strands where more than one PWR ratio is advertised by "D".

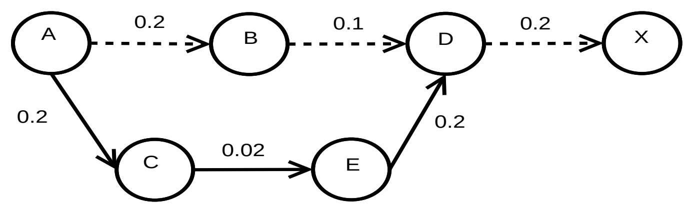

Figure 9. Low-power path derived using the algorithm that uses low value ingress link but through the same AS.

of the Internet receive the attribute and the time it originated. This however cannot be avoided as with other attributes and metrics.

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 neighbouring 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 an appropriate 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 neighbouring AS, and later propagated 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 to be 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 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.

Multiple ASBRs advertising differing PWR ratio can lead to AS that have low PWR ratio through an ingress link and not through other. Consider the case of multiple ASBRs that belong to the same AS, advertising differing PWR ratios. This could lead to power values that belong to different classes with 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. As an example, 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.

V. Conclusion and Future Work

In this paper, we have proposed a scheme for reducing the power consumption of the Internet using collaborative effort between AS. The topology of the Internet is depicted as a graph 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 determines the low-power consuming path between AS and routes data packets from head-end to tail-end. Explicit routing is handled through the use of TE-LSPs. Since using CSPF can be time consuming a heuristic algorithm to derive the low-power paths using graph-labelling is 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.

The scheme proposed in this paper assumes that the PWR ratio information is reliable. It is possible that ISPs could fake the PWR ratio information. However, ISPs usually have service level agreements (SLAs) for carrying traffic. One method is to link up each ISP with a power application level gateway to ensure that proper ratios are advertised. This could be mandated at least amongst the cooperating ISPs. Further 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. Our future work will quantify and analyse these issues.

Acknowledgments

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.

References

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

[2] 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, doi: 10.1109/COINA-COFT.2007.4519173.

[3] A. Bianzino, C. Chaudet, D. Rossi and J. Rougier, "A survey of green networking research", IEEE Communications and Surveys Tutorials, preprint, pp. 1–18, doi: 10.1109/SURV.2011.113010.00106.

[4] F. Blanquicet and K. Christensen, "Managing energy use in a network with a new SNMP power state MIB", IEEE Conference on Local Computer Networks, October 2008, pp. 509–511, doi: 10.1109/LCN.2008.4664214.

[5] R. Bolla, R. Bruschi, A. Cianfrani and M. Listani, "Enabling backbone networks to sleep", IEEE Network, vol. 25, no. 2, March/April 2011, pp. 26–31, doi: 10.1109/MNET.2011.5730525.

[6] R. Bolla, R. Bruschi, F. Davoli and F. Cucchietti, "Energy efficiency in the future Internet: A survey of existing approaches and trends in energy-aware fixed network infrastructures", IEEE Communications Surveys and Tutorials, vol. 13, no. 2, second quarter 2011, pp. 223–244, doi: 10.1109/SURV.2011.071410.00073.

[7] J. Chabarek, J. Sommers, P. Bardford, C. Estan, D. Tsiang and S. Wright, "Power awareness in network design and routing", Proceedings of the IEEE INFOCOM 2008, April 2008, pp. 457–465, doi: 10.1109/INFOCOM.2008.93.

[8] H. Chang, R. Govindan, S. Jamin, S. J. Shenker and W. Willinger, "Towards capturing representative AS-level Internet topologies", Computer Networks, vol. 44, April 2004, pp. 737–755, doi: 10.1016/j.comnet.2003.03.001.

[9] A. Cianfrani, V. Eramo, M. Listanti and M. Polverini, "An OSPF enhancement for energy saving in IP networks", Computer Communications Workshops, INFOCOM 2011, April 2011, pp. 325–330, doi: 10.1109/INFCOMW.2011.5928832.

[10] 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, doi: 10.1109/TNET.2009.2034143.

[11] R. Oliveira, D. Pei, W. Willinger, B. Zhang and L. Zhang, "The (in)completeness of the observed Internet AS-level structure", IEEE/ACM transactions on Networks, vol. 18, no. 1, February 2010, pp. 109-122, doi: 10.1109/TNET.2009.2020798.

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

[13] M. Xia, M. Tornatore, Y. Zhang, P. Chowdhury, C. Martel and B. Mukherjee, "Greening the optical backbone network: A traffic engineering approach", IEEE ICC Proceedings, May 2010, pp. 1–5, doi: 10.1109/ICC.2010.5502228.