# Using BGP to Reduce Power Consumption in Core and Edge Networks: A Metric-Based Approach

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*—Power reduction methods at the device and the network levels in the Internet continue to attract attention. At the device level, power reduction is usually achieved by using low-power consuming devices or by switching off unused components. At the network level; re-engineering, reconfiguring, and re-routing of data packets can help in reducing power consumption. In this paper, we present a metric-based approach to route data packets through a low-power path from a source to a destination at the level of Autonomous Systems (AS). We propose that the Border Gateway Protocol (BGP) exchange a new attribute; namely, *consumed-power to available-bandwidth* of an AS with neighbouring AS. Using this proposed metric, the AS Border Routers can readily identify the low-power path from a source to a destination. We propose appropriate modifications to the BGPs path selection algorithm to include the low-power path criteria. We consider the effects that the proposed approach has on two parameters: (a) power reduction achieved as compared to the shortest path routing, and (b) the increase in path length from source to destination. The increase in the number of hops would be a consequence of re-routing through low-power AS. Simulations show that there could be significant gains in power reduction, if an increase in the number of hops is acceptable. Our work suggests that this trade off merits consideration as the power consumption of an AS is significant.**

***Index Terms*—Autonomous Systems, Border Gateway Protocol, Low-power Paths, Traffic Engineering.**

## I. Introduction

Power reduction methods that aim to reduce energy consumption of the Internet have evoked much interest. In [1], a metric-based method for reducing the power consumption of the core and edge networks was proposed. This power reduction problem assumes significance as estimates predict a $300\%$ increase, when access speeds move from $10$ Mbps to $100$ Mbps [2]. Numerous approaches have been proposed to reduce the power consumption ranging from designing low-power routers and switches, to 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 [4], it is shown that the link power consumption can vary by $20$ Watts from the base power, between idle and traffic scenarios. Hence, the authors suggest fully utilizing the line-card. The idea is that 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 [5]. Some variants of TCAMs have been proposed for high speed lines with reduced power consumption [6]. But these schemes cannot scale forever. For some modeling work associated with buffer sizes, which can also lead to a reduction in power at the router architecture level, see [7], [8], and references therein.

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 [10], 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 $4$ Watts 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. A detailed review on energy efficiency of the Internet is given in [11].

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 [12], an approach for reducing intra-AS power consumption for optical networks using Dijkstra'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 an intra-AS scenario. Traffic is then re-routed through the low-power optimized links.

The following issues exist in power reduction schemes today.

a A common method for reducing power consumption in the Internet is to switch unused devices and links to sleep state. Data packets are then routed through the functional links. This method is very localized and does not consider the power increase in the adjacent device that carries the extra traffic. Further this solution

is dependent on the underlying technology used by the devices. Also service level agreements between Internet Service Providers (ISPs) may be such that switching off the links may not be permissible.

b Power reduction schemes should not operate in isolation. They must be hierarchical so that they are applicable at various Internet hierarchies such as within the enterprise, or AS, and between AS. Further, there must not be any large variation in the algorithms implemented at the various levels of hierarchy. If multiple schemes are implemented at various levels of the hierarchy, then a way to coordinate these schemes become essential.

c Distributed solutions for power reduction have been used in adhoc wireless networks [13]. Such schemes may not be extensible to large networks such as the Internet. Further any proposed scheme must be extensible to multicast networks as well and not be limited to unicast.

Some of these drawbacks were addressed in our earlier scheme which was applicable for Multi-Protocol Label Switching (MPLS)-based networks [1]. MPLS label switched paths that traverse multiple AS carry traffic from a head-end to a tail-end AS, use Border Gateway Protocol (BGP) for exchanging routing and topology related information. In [1], the low-power path was detected by identifying the topology of the Internet. This topology at the AS level was obtained using the method presented in [14], where one of the attributes of BGP, AS-PATH-INFO, was used. The Constrained Shortest Path First algorithm (CSPF) uses this AS level topology with *consumed-power to available-bandwidth* (PWR) metric as a constraint, to determine the low-power path from the head-end to the tail-end. The PWR metric can be exchanged among the collaborating AS using BGP. It was shown that 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 Path (TE-LSP) that span multiple AS. However, this method has communication overhead in order to setup the path. These overheads occur in the form of information exchange between the entities in the network.

In order to avoid this, we propose modifications to the BGP path selection algorithm. This reduces the communication overhead associated with respect to setting up of the path. We introduce a new path selection rule to ensure that routing paths are established based on PWR metric by BGP rather than using inter-AS TE LSP. Simulations show that the PWR metric-based algorithms can lead to a power reduction which is as high as $70\%$ over the conventional CSPF hop based variant. The power reduction obtained depends on the connectivity of the topology as well as the PWR metric distribution. There could be up to a $50\%$ increase in the number of hops when compared with the shortest path algorithm. It has been suggested that for Internet Protocol (IP)-based networks such increase in hops may not have much impact in performance at the application layers [4].

The rest of the paper is organized in the following manner. In Section II, an overview of the BGP routing protocol as well as the inter-AS TE-LSP based algorithm is presented. Section III addresses ways to reduce the communication time complexity by proposing a method for establishing low-power paths using BGP path selection. Simulations are discussed in Section IV. In Section V, a brief discussion on the implementation and emulation using OpenFlow and Quagga is presented. A discussion on the comparison with our previously proposed implementation is presented in Section VI. We outline our contributions and highlight avenues for future research in Section VII.

## II. Preliminaries

In this section, we present an introduction to the BGP protocol and the inter-AS TE-LSP based method for inter-AS power reduction.

### A. Border Gateway Protocol

BGP [15] performs routing between multiple AS by exchanging routing and reachability information with other systems implementing BGP. BGP installs routing tables using a path selection algorithm. Routing information exchanges happen between multiple AS. This is classified under Exterior Border Gateway Protocol (eBGP). Internal BGP (iBGP) peering is used between the border routers in an AS and if necessary, between the core routers as well. Internal BGP expects a mesh topology. As such a network topology can become unmanageable due to scalability, route reflectors that re-advertise only the best path information is used to convey information to other iBGP routers.

BGP's routing decision is based on various static and dynamic parameters. Some examples for static parameters that affect routing decisions include multi-exit discriminator (MED) and local preference (LOCAL_PREF) values. Routing through oldest paths and AS-Path lengths are some examples for dynamic parameters. For a detailed discussion refer to [16], [17].

### B. Inter-AS TE LSP power reduction using BGP

In our previous work, we presented a methodology for addressing the power reduction problem in the core and edge networks using BGP. The methodology was divided into four parts:

1) constructing the topology of the AS by a device,
2) assigning the PWR metric to the links connecting the AS,
3) calculating the low-power paths in the AS topology, and
4) establishing the path from source to the destination using traffic-engineering techniques.

We now briefly review the algorithms and discuss the computation and communication time complexity issues.

*1) Constructing the 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, we use the approach from [14]. In this approach

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 PWR metric of the AS, as seen from an AS Border Router (ASBR). Since there can be more than one ASBR associated with an AS, a vertex can have more than one PWR metric. Note that the ASBRs act as an entry point to the AS. Each of the PWR metrics for a vertex are assigned to the ingress links of the ASBRs. 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 had $46$ nodes with $15\%$ connectivity in the topology. We define connectivity as a percentage of links present in the topology when compared with a complete graph of $N$ nodes, which is $\frac{N(N-1)}{2}$.

*2) PWR metric calculation:* The numerator of the PWR metric 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 technique. The idea is to obtain the consumed-power of the AS which is the averaged consumed-power for all the routers within an AS. This value is divided by the maximum available-bandwidth at each of the ASBRs 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 available-bandwidth in the ASBRs egress links. Note that Simple Network Management Protocol (SNMP) can also be used to extract this power information [18] offline.

The highest available bandwidth amongst the ASBRs egress links is used as the denominator in the PWR metric computation. Once the requested bandwidth is available, then consumed power plays a major role in determining the path from the source to the destination. PWR metric is used as a mapping function for each of the ingress link of the ASBR of an AS. This metric is then advertised to the other neighboring AS through the control plane using BGP extensions. BGP ensures that the information is percolated to other AS. On the receipt of these PWR metrics by the AS at far-end of the Internet, the overall AS level topology can be constructed. Note that 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 are obtained by using such a mapping function on the ingress links.

*3) Low-power path detection:* The algorithm consists of two sub-algorithms: each one executed by the ASBRs and the Path Computation Elements (PCEs) in the network in their respective AS. PCEs have been proposed by the Internet Engineering Task Force 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 and Algorithm 2, respectively.

In Algorithm 1, parallel process $1$ (steps $4-10$) is used to exchange the PWR metric information. Parallel process $2$ (steps $11-16$) handles the TE-LSPs. Algorithm 2 calculates the low-power path from the head-end to the tail-end and sends this path information to the head-end AS.

**Algorithm 1** ASBR low-power path algorithm

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

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

**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: Calculate the shortest paths from the head-end to the tail-end using CSPF with PWR as a metric;
3: if no path available then
4:     Signal error;
5: end if
6: if path exists then
7:     Send explicit path to head-end to construct path;
8: end if
9: Continue passively listening to BGP updates to update T=(AS, E, f);
10: End
```

*4) Path establishment:* Using the PWR metric the low-power path is obtained by applying the CSPF algorithm. For example, in Figure 2, the path $(A, B, D, G, H, X)$ is power efficient as the summation of the PWR metric in this path is minimum when compared with other paths in the graph topology. Of course, 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

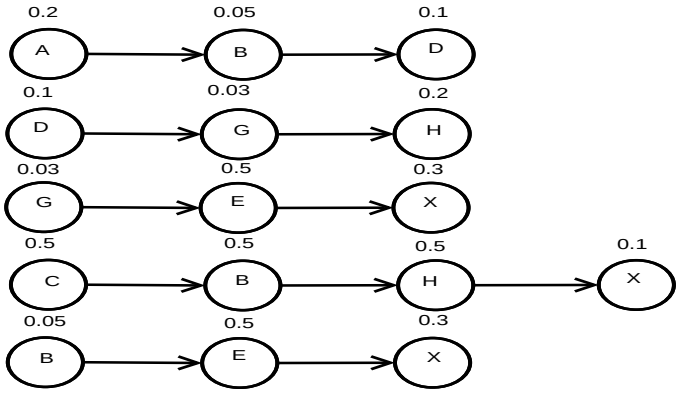


Fig. 1. Strands obtained from BGP updates, vertices $A, B, C, D$ and $G$ are the head-end AS; $D, H$ and $X$ are the tail-end AS. The vertex weights represent the PWR metric of an AS, and the link direction shows the next AS hop. ASBRs present the topology to the PCE.

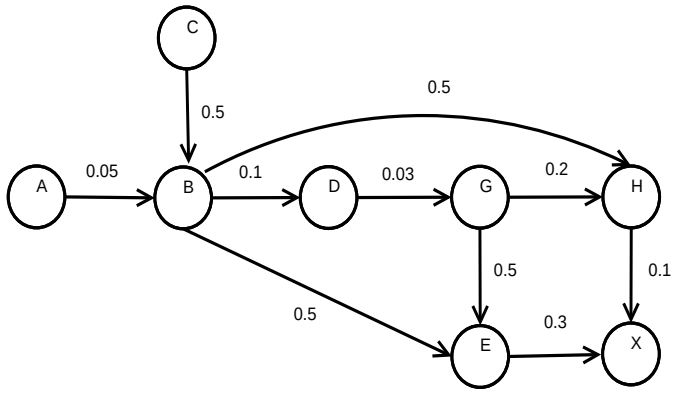


Fig. 2. Strands combined to get the Internet topology. The PWR metric is mapped to the ingress link of the ASBR. CSPF algorithm is run on this topology to detect the low-power path by the PCE.

is established. 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 fall back to the conventional 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.

### C. Time complexity

In protocol related algorithms two issues are of interest namely: communication and computation time complexity.

*1) Communication time complexity:* The communication time complexity involved in the algorithm includes

- monitoring the BGP packets to discover the topology based on AS-PATH-INFO attribute,
- exchanging PWR metric between the neighboring AS, and
- using inter-AS TE LSP to construct the path from the head-end to the tail-end AS.

Monitoring the BGP packets takes $O(1)$, a constant time. Exchanging PWR metric between the neighbors occurs in a distributed manner and hence takes $O(1)$, a constant time. The construction of the path takes $O(N)$ where $N$ is the diameter of the network topology. Hence, the computational complexity is $O(N)$.

*2) Computational time complexity:* The algorithm for forming the topology from AS strands is dominated by the number of links. In the case of dense connectivity the computational time complexity is $O(|V| + |E|) \approx O(|E|)$, where $|E|$ is the number of edges and $|V|$ is the number of nodes or vertices. The computational time complexity of the low-power path algorithm is dominated by the Dijkstra's algorithm. In this case, instead of hops or any other metric we use the PWR metric in the Dijkstra's shortest path algorithm. Hence, the computational time complexity is bounded by Dijkstra's shortest path algorithm which is $O(|E| \log |V|)$.

The algorithm discussed above can be implemented "offline". The topology information could be extracted by passive monitoring, PWR metric information can be obtained using SNMP, and the low-power path can be calculated using a separate offline system. The paths can then be established by installing routing tables remotely. This offline implementation has certain drawbacks. We list the drawbacks and some possible solutions to overcome them.

1) 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.
2) The topology information obtained using the BGP strands might be incomplete. For a detailed discussion on completeness of Internet topology using BGP refer to [19], [20]. In addition to the BGP based algorithm, any other algorithm such as SNMP based Topology discovery could also be used to enhance connectivity as well as discover new nodes. But this increases the communication time complexity.
3) The PCEs usually use modified Dijkstra's shortest path algorithm and not a distributed algorithm. The algorithm can be speeded up using the graph-labeling method discussed in [1].
4) The algorithm uses RSVP and inter-AS TE LSP to establish the path which leads to communication overhead.

As we can see, too many information exchanges happen to implement the low-power path selection process. It would be of interest to see whether we could reduce the communication time complexity issues. Note that the computational time complexity is still bounded by the shortest path algorithm. In the next section, we explain the BGP path selection algorithm which overcome these issues.

## III. BGP LOW-POWER PATH SELECTION

Before we study the proposed changes to the BGP based path selection algorithm, we will review the current algorithm

discussed in [16]. These algorithms are executed by the ASBRs and the core routers.

### A. BGP path selection

In the BGP algorithm [15], each entity exchanges the best route to a given destination with other connected entities. Therefore, the BGP protocol is effectively a distributive method for generating routing information and there is no need to explicitly discover the topology. Of course this means that the information obtained from the neighboring entities must be reliable which is the case in the Internet. Such a distributed BGP algorithm exchanges prefixes and their next hops after going through the best path selection steps. Hence, there is a need to compare and choose the best route to add to the IP routing table which is used for routing the data packets. For this process to take effect, BGP uses about thirteen different rules to choose the path [16], [17]. We add the PWR metric-based low-power path selection as another rule.

The algorithm works as follows: BGP assigns the first valid path as the current best path based on the paths it received from the neighboring entities. BGP then compares the best path with the next path in the list, until BGP reaches the end of the list of valid paths. The rules that are used to determine the best path are given briefly in Algorithm 3.

There are some exception conditions in some of the steps; for details refer [16]. We now modify the BGP path selection algorithm functionality by including the low-power path PWR metric-based calculation. This involves adding Algorithm 1 as a subroutine to the BGP path selection criteria after line 4 and expanding line 5 of Algorithm 3, where we select the shortest path only if a low-power path is not available. The following conditions are considered in the PWR metric-based low-power path selection (see Algorithm 4).

1) If the PWR metric is not available for a link, we drop all paths using the link (steps $20-22$).
2) If the PWR metric-based detection is not enabled in even a single entity that uses BGP we do not execute this algorithm (steps $5-6$).
3) If there is only one AS PATH then we skip applying this algorithm (default action).
4) If there are multiple paths to the tail-end then we choose the one with the least sum of PWR metric to the tail-end (step $14$).
5) If multiple path exists with the same PWR sum, then we choose all the paths and give it to the path selection algorithm (steps $13-18$).
6) If there are no PWR metric-based paths we fall back to the shortest path algorithm (default action).

The detailed changes are given as Algorithm 4. Algorithm 3 is now complete with the inclusion of the low-power path selection process using the PWR metric.

Note that this method involves changes to the BGP path selection algorithm and hence all the devices involved in exchanging BGP routes must implement this method. Therefore, this method cannot be implemented offline. We will refer to this as "online" implementation.

**Algorithm 3** Abridged BGP algorithm

**Require:** Topology information related with BGP

1: Begin
2: Prefer the path with the highest WEIGHT a locally configured parameter for a router.
3: Prefer the path with the highest LOCAL_PREF value, a value configured for local preference.
4: Prefer the path that was locally originated via a network or aggregate BGP subcommand or through redistribution from an Interior Gateway Protocol.
5: Prefer the path with the shortest AS_PATH, the shortest path from source to destination.
6: Prefer the path with the lowest origin type (Exterior Gateway Protocol paths preferred over Interior Gateway Protocol paths).
7: Prefer the path with the lowest multi-exit discriminator (MED). This parameter is used when there are multiple paths to a destination.
8: Prefer external BGP over internal BGP paths.
9: **if** bestpath is selected **then**
10: go to MULTIPATH;
11: **end if**
12: Prefer the path with the lowest IGP metric to the BGP next hop.
13: MULTIPATH: Determine if multiple paths require installation in the routing table for BGP Multipath.
14: **if** best path selected **then**
15: exit with the best path.
16: **end if**
17: When both paths are external, prefer the path that was received first (the oldest path).
18: Prefer the route that comes from the BGP router with the lowest router ID.
19: If the originator or router ID is the same for multiple paths, prefer the path with the minimum cluster list length. The router ID is the highest IP address on the router, with preference given to loopback addresses.
20: Prefer the path that comes from the lowest neighbor address.
21: End

### B. Time complexity

We now discuss the communication and computational complexity of the proposed algorithm.

*1) Communication time complexity:* Topology discovery is not needed in this algorithm as BGP exchanges best routes with the neighboring entities. This removes the need for using TE-LSPs to establish the path from the head-end to the tail-end AS as well. Of course, traffic engineering techniques can still be enforced. There is no additional communication overhead other than the addition of a new BGP attribute to the BGP protocol. Therefore, the total communication complexity is bounded by a constant, $O(1)$.

**Algorithm 4** Modified BGP path selection algorithm

**Require:** BGP path selection algorithm

```
Begin
if ROUTER is configured with BGP then
    Execute Step 2, 3 and 4 of Algorithm 3
end if
if there is no PWR metric-based path selection then
    Goto Step 5 of Algorithm 3;
else
    if (there are multiple AS_PATHS) AND (PWR metrics) then
        Calculate the sum of PWRs in the paths.
    else
        Ignore paths that have no PWR metrics.
    end if
    if there exists multiple sum of PWRs as there is more than one path then
        Choose the AS_PATHS with the least PWR metric sum.
        if multiple least PWR metric sum are equal then
            Choose all the AS_PATHS;
            Goto Step 6 of Algorithm 3;
        end if
    else
        if there exist no PWR_SUM because of exclusion then
            Goto Step 5 of Algorithm 3;
        end if
    end if
end if
Goto Step 6 of Algorithm 3;
End
```

*2) Computational time complexity:* The computational time complexity of the BGP path selection algorithm is bounded by the calculation of the path with the smallest sum of PWR metric, if multiple paths exist. In the worst case, we might have to apply Dijkstra's shortest path algorithm with PWR metric on the topology learned by the ASBRs. Therefore, the computational time complexity is still bounded by that of Dijkstra's algorithm which is $O(|E| \log |V|)$, with $|E|$ and $|V|$ representing the number of edges and nodes, respectively. By using the proposed algorithm, we overcome the draw backs of the inter-AS TE-LSP based low-power path algorithm. Note that the path selection is done by the ASBR and there is also no need for the use of PCE in this method.

We conducted simulations using the offline PWR based method to study the possible power reduction.

## IV. Simulations

The simulations involved creating various graph topologies for a given connectivity. For large values of vertices $V$ GNU scientific library based simulation was performed [21]. We assumed a uniform link distribution between the nodes. Uniform distribution was used to ensure that there were minimal number of disconnected components under low connectivity. On this topology, PWR metric values based on uniform and exponential distribution were assigned to the links. The experiments were repeated for different set of values of distributions. We considered about $100$ topologies for each connectivity ranging from $20\%$ to $90\%$. Any graph topology that was disconnected was dropped from the study. We used the Dijkstra's algorithm for finding the low-power as well as shortest paths. The simulation parameters are given in Table I.

TABLE I
Simulation parameters and their values

| *Parameter* | *Value* |
|---|---|
| Topology size | $100, 10000,$ 1million nodes |
| Connectivity | $25 - 95\%$, step size $5\%$ |
| Low-power nodes | Uniform, Exponential ($\lambda = 0.25$) |
| Network types | 100 topologies for each connectivity |

Two important parameters were monitored: the increase in the number of hops and the comparative power reduction possible by opting for the low-power path algorithm. For each source-destination pair in the topology, we compared the power reduction obtained by using low-power paths with that of the conventional hop based shortest path metric. The PWR metric can vary dynamically over a time period which also means that the low-power paths can vary for a given connectivity. Therefore, we also monitored the power-reduction and hop variations for a connectivity of $70\%$. The graphs presented are for node size $100$. For $70\%$ connectivity we show $30$ randomly chosen sample topology out of the $100$ topologies that were used.

### A. Uniform distribution of PWR metric

The graphs shown in Figure 3 for uniform distribution of PWR metric depicts the power reduction and hop increase for various connectivity size. We see that the minimum power reduction that can be achieve is around $10\%$ and this increases almost linearly with connectivity. The hops can also increase by up to $50\%$. High values of power reduction were possible as there are equal number of links with high and low PWR metric under this distribution. It can be seen that the average number of hops increases with more connectivity. This is because more low-power links also increase under such circumstance and the algorithm prefers routing through such low-power links.

### B. Exponential distribution of PWR metric

In this case, the topology had more low-power links. From Figure 4, it can be seen that the average power reduction as well as hop increase could be as high as $65\%$. This is possible as the network topology is biased towards low-power links. The hops also increase as the proposed method tries to use all the low-power links for establishing routing information. The simulations for the two distributions establish that the algorithm uses low-power paths to route data packets. It should be noted that the PWR metric uses the power consumption of

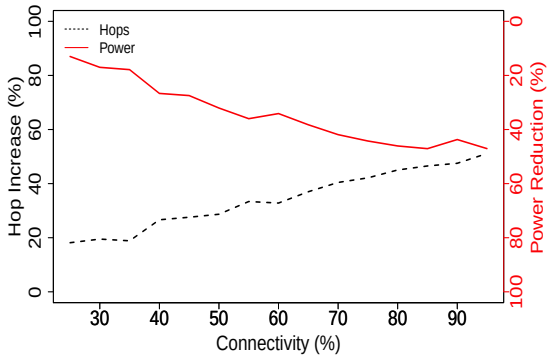


Fig. 3. Uniform distribution of power links

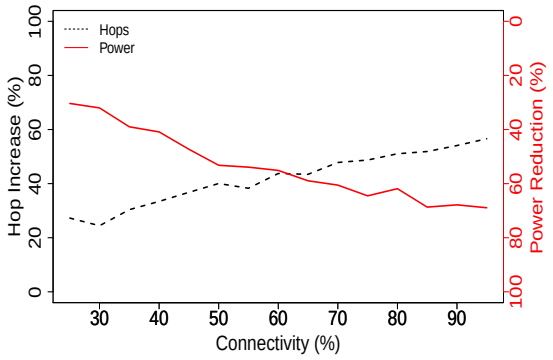


Fig. 4. Exponential distribution of power links

the AS. Even though the simulations can be considered very optimistic, typically AS consume Mega-Watts of power [22], [23]. Therefore, even a $1\%$ reduction in power of an AS can result in significant savings for ISPs.

These results also suggest that after a particular value of connectivity size, increase in hops does not return much benefit with respect to power reduction. Therefore, it is interesting to study the behavior of the algorithm under a given connectivity value. Such a study will help to understand the dynamics of the network. Note that the PWR metric will also fluctuate over time and hence the paths can dynamically change.

### C. Role of connectivity

We fixed a connectivity of $70\%$ and studied the network for power reduction and hop increase using both the distributions (see Figures 5, 6). Results indicate that the graph topology plays a major role in power reduction. For both the distributions, the power reduction as well as increase in hops is bounded by a range which is dictated by the connectivity. The PWR metric will vary over a period of time. Therefore, the algorithm also reschedules the routing information based on the low-power values. Each trial can even be considered as the network topology at a time instant. The average power reduction remained quite high in both the cases.

We now discuss some implementation issues of the proposed online algorithm.

## V. Implementation

In this section, we present notes on feasibility of implementation in a live network. We also briefly discuss implementation work based on OpenFlow [24] for offline implementation and Quagga [25] based online implementation.

### A. Feasibility of implementation

First, the requested bandwidth should be available on the low-power path. This can be taken care using TE methods. Second, there is a reliable flooding process that gets triggered when updates about the change in PWR metric 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 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 metric as well. The proposed PWR metric is advertised to the neighboring AS and the information percolated to all the AS, in a AS-PATH-POWER-METRIC attribute. This attribute is discussed in the Appendix. 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 metric 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 neighboring 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 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 metric has changed in the originating AS.

The power consumed by each router may fluctuate over short time intervals. This can occur if the data packets are re-routed. In this case a low-power path might start consuming higher power and advertise a higher PWR metric. It is possible that the routes can flap due to PWR metric changes. In order to dampen these fluctuations, power can be measured when falling within suitable intervals as opposed to a discrete quantity. This method of power measurement reduces the frequency of triggered updates from the routers due to power change. This can sometimes affect the network performance. This situation must also be addressed while using PWR metric-

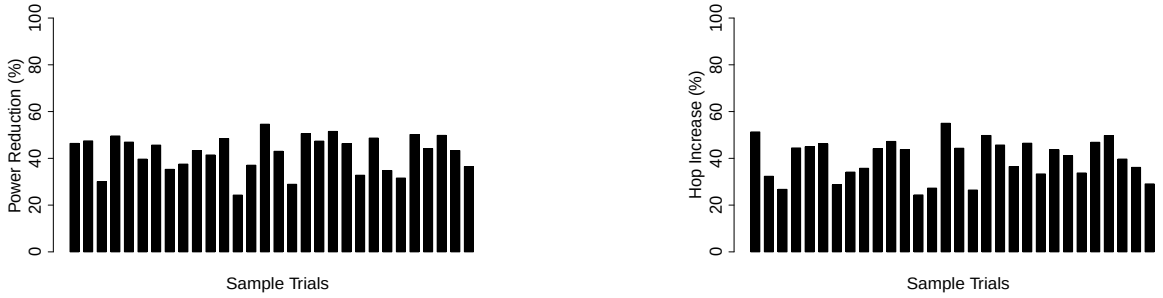


Fig. 5. Power reduction and hop increase for *uniform distribution* of power values with 70% connectivity.

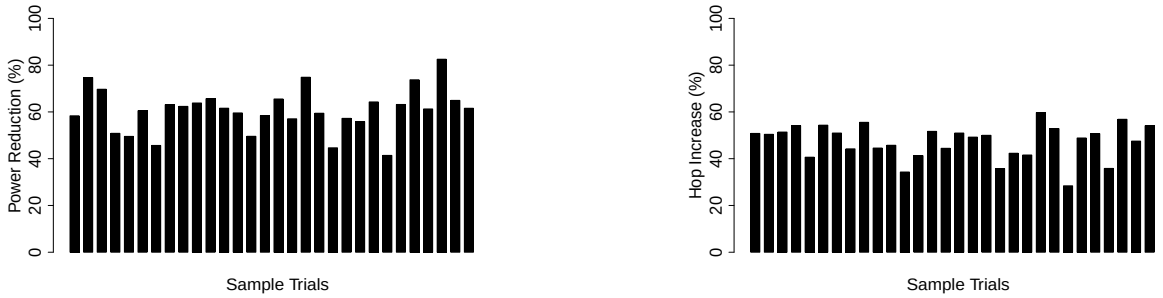


Fig. 6. Power reduction and hop increase for *exponential distribution* of power values with 70% connectivity.

based methods in the network.

Multiple ASBRs advertising differing PWR metric can lead to AS that have low PWR metric through an ingress link and not through other. Consider the case of multiple ASBRs that belong to the same AS, advertising differing PWR metrics. This could lead to power values that belong to different classes with intervening classes in between. These advertised PWR metrics 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 7 which shows a set of strands that derive a topology as in Figure 8. Here, $D$ is reachable via two paths but the PWR metrics 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.

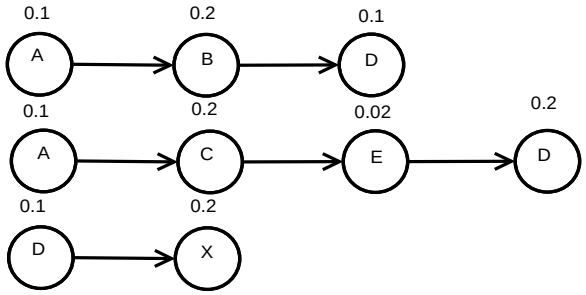


Fig. 7. Example of strands where more than one PWR metric is advertised by $D$.

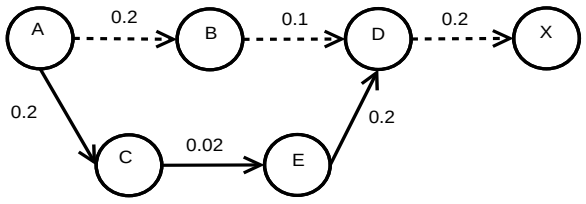


Fig. 8. Low-power path derived using the algorithm that uses low value ingress link but through the same AS.

### *B. OpenFlow implementation*

Since we did not have access to a live network, we emulate the algorithms using a simple offline implementation based on OpenFlow [24]. OpenFlow was designed to run experimental protocols on the campus network for research purposes. Many of the vendor devices support OpenFlow as a part of their capability. The control flow part of the router/switch is handled by a OpenFlow controller while the data path still resides at the router. OpenFlow devices and the OpenFlow controller communicate with each other using OpenFlow protocol [24].

In our implementation, the router/switch are treated as Autonomous Systems. These devices present information about

their consumed-power to the centralized controller. To simplify the implementation, we assume that enough bandwidth is available at these routers. This is a realistic assumption as enough bandwidth is usually available in the core routers. The centralized controller determines the topology of the network based on the connectivity information obtained from the routers. Using this topology as well as the consumed-power information, the centralized controller updates the routing tables so that data packets traverse through low-power paths.

### C. Quagga based implementation

As a part of online implementation strategy, Quagga based implementation is studied [25]. The BGP daemon is modified and routes are formulated based on the low-power path criteria. The power information is obtained through the use of an experimental MIB included in the Quagga based Linux routers. At a later stage we plan to incorporate the AS-POWER-PATH-METRIC discussed earlier.

## VI. Discussion

In this section, we first compare the offline and the online approaches. We then outline Quality of Service (QoS) aspects that need to be considered to compensate for any increase in the number of hops when low-power paths are chosen.

### A. Comparison of offline and online approaches

In the offline approach, TE-LSPs are needed to establish the path between the source and the destination. In contrast, in the online approach, separate TE based label switching can be completely avoided. The routing table is generated using the BGP algorithm itself and hence the overhead for establishing the paths is reduced considerably. To incorporate the online approach the BGP path selection algorithm has to be modified in all the core and edge routers implementing this scheme. The online approach is considerably faster than the offline approach with a trade off in the implementation complexity.

### B. Latency in the network

Finding low-power paths in the graph topology might lead to an increase in the number of hops between a source and a destination. An increase in the number of hops could lead to an increase in queuing and propagation delay. Propagation delay is unavoidable as it depends on the transmission medium. Queuing delay is introduced rather naturally due to the store and forward design of Internet routers and also as a consequence of the design of the flow control methods implemented in transport protocols. Latency is a key QoS metric, and thus minimising end-to-end delay is an important network engineering task. To compound the problem, router buffers are currently sized based on an out-dated bandwidth-delay product rule which was intended to maintain full link utilisation. There are two options to reduce queuing delays: either reduce the buffer sizes dramatically [7], or make judicious use of feedback, from queues, to design better queue management policies [8]. A low-latency network thus improves QoS and could possibly enable the deployment of power saving methods which might require an increase in the number of end-to-end hops. Internet QoS is an area of active research among the communication networks community.

## VII. Outlook

We propose a method, which employs a collaborative approach between AS, to reduce power consumption by using the Consumed-Power to Available-Bandwidth (PWR) metric.

### A. Contributions

In our previous work [1], the AS topology was represented as a graph using the strands obtained from the AS-PATH-INFO attribute of the BGP updates. The CSPF algorithm was run on this topology by using the PWR metric as an additional constraint. The PWR metric is advertised through the ingress links of the ASBRs associated with AS using BGP updates. Inter-AS Traffic Engineering Label Switched Paths were used to route the data packets from the head-end to the tail-end. As using CSPF can be time consuming a heuristic algorithm to derive the low-power paths using graph-labelling was proposed. The communication time complexity associated with information exchanges in this method is high.

In order to reduce this complexity, in this paper we proposed that the BGP path selection algorithm be used to determine the low-power consuming paths between AS using the PWR metric. To study the performance and viability of using PWR metric-based methods, we conducted simulations on various topologies with different PWR metric distributions. The distributions used were the uniform and exponential distributions, and the results were especially encouraging: there was a substantial gain in power reduction where the tradeoff was an increase in the number of hops. We also briefly discussed emulating these schemes with OpenFlow and Quagga based BGP. Given the current power consumed by the AS, reduction in power savings could be rather beneficial to the ISPs.

### B. Avenues for future work

The methods proposed in this paper assume that the PWR metric information is reliable. An erroneous metric information can be a cause of concern. 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 metrics are advertised. This could be mandated at least amongst the cooperating ISPs.

It would be of interest to study whether the conceptual methods used at inter-AS level can be employed to inter-Area based topology. It is also natural to extend the study for multicast traffic. It would certainly be interesting to perform an evaluation of the proposed methods on a range of topologies and PWR metric distributions. Our work focused on the core and access networks that use BGP as the routing protocol. A study on extending these methods to other access networks implementing wireless connections would be useful. We have not considered the role of different traffic distributions on

power consumption. A practical study could be conducted on a live AS topology.

It has recently been highlighted that queuing delay in the Internet is on the rise [26]. The proposed scheme for power reduction would lead to an increase in the number of hops. Thus significant queuing delays at each hop would negatively impact QoS if the number of hops, between source and destination, are increased. Given the potential benefits for power reduction it would be imperative to investigate the design of queue management schemes to ensure a low latency network. Some work in this direction has already been started [8].

## ACKNOWLEDGMENT

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 thank Prof. Kamakoti for allowing the use the RISE lab facilities for our simulations. We appreciate the helpful suggestions by Fabrice Saffre and Hanno Hildmann on the presentation.

## REFERENCES

[1] S. Raman, B. Venkat, and G. Raina, *Reducing power consumption using the Border Gateway Protocol*, Proc. of the Second International Conference on Smart Grids, Green communications and IT Energy-aware Technologies, Energy 2012, March 2012, pp. 83–89, ISBN: 978-1-61208-189-2.

[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/COINACOFT.2007.4519173.

[3] A. P. Bianzino, C. Chaudet, D. Rossi, and J. Rougier, *A survey of green networking research*, IEEE Communications and Surveys Tutorials, preprint, 2011, pp. 1–18, doi: 10.1109/SURV.2011.113010.00106.

[4] J. Chabarek, J. Sommers, P. Bardford, 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, doi: 10.1109/INFOCOM.2008.93.

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

[6] W. Lu and S. Sahni, *Low-power TCAMs for very large forwarding tables*, IEEE/ACM Transactions on Computer Networks, vol. 18, no. 3, June 2010, pp. 948–959, doi: 10.1109/TNET.2009.2034143.

[7] G. Raina, D. Towsley, and D. Wischik, *Part II: control theory for buffer sizing*, ACM SIGCOMM Computer Communications Review, vol. 35, no. 3, July 2005, pp. 79–82, doi: 10.1145/1070873.1070885.

[8] S. Raman, S. Jain, and G. Raina, *Feedback, transport layer protocols and buffer sizing*, Proc. of the Eleventh International Conference on Networks, February 2012, pp. 125–131, ISBN:978-1-61208-183-0.

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

[10] 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.

[11] 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.

[12] 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.

[13] G. Y. Li et.al., *Energy-efficient wireless communications: tutorial, survey, and open issues*, IEEE Wireless Communications, vol. 18, no. 6, 2011, pp. 28–35, doi: 10.1109/MWC.2011.6108331.

[14] B. Venkat, A. V. Rajagopalan, and B. Bhikkaji, *Constructing disjoint and partially disjoint InterAS TE-LSPs*, USPTO Patent 7751318, Cisco Systems, 2010.

[15] Y. Rekhter and T. Li,A border gateway protocol 4 (BGP-4). *http://tools.ietf.org/html/rfc4271*. [Accessed: December 6, 2012].

[16] BGP path selection algorithm, *http://www.cisco.com/en/US/tech/tk365/technologies_tech_note09186a0080094431.shtml*, last accessed [December 6, 2012].

[17] BGP path selection algorithm, *http://www.juniper.net/techpubs/en_US/junos11.4/topics/reference/general/routing-ptotocols-address-representation.html*, last accessed [December 6, 2012].

[18] 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.

[19] 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.

[20] 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.

[21] M. Galassi et.al., *GNU Scientific Library Reference Manual, 3rd Edition, ISBN: 0954612078*, http://www.gnu.org/software/gsl/manual/html_node/, last accessed [December 6, 2012].

[22] K. Hinton, J. Baliga, M. Z. Feng, R. W. A. Ayre, and R. S. Tucker *Power consumption and energy efficiency in the internet*, IEEE Network, vol. 25, no. 2, March-April 2011, pp. 6–12, doi: 10.1109/MNET.2011.5730522.

[23] A. P. Bianzino, L. Chiaraviglio, M. Mellia, and J. L. Rougier, *GRiDA: Green distributed algorithm for energy-efficient IP backbone networks*, Computer Networks, vol. 56, no. 14, 2012, pp. 3219–3232, doi: 10.1016/j.comnet.2012.06.011.

[24] N. McKeown, *OpenFlow: enabling innovation in campus networks*, ACM SIGCOMM Computer Communication Review, vol. 38, no. 2, 2008, pp. 69–74.

[25] O. Bonaventure, *Software tools for networking*, IEEE Network, vol. 18, no. 6, 2004, pp. 4–5.

[26] K. Nichols and V. Jacobson, *Controlling queue delay*, Communications of the ACM, vol. 55, no. 7, 2012, pp. 42–50, doi: doi.acm.org/10.1145/2209249.2209264.

## APPENDIX

The proposed AS-POWER-PATH-METRIC attribute is shown in Figure 9. Since the updates can be triggered quite frequently, sequence numbers are needed. The rest of the fields are needed to exchange the PWR information and are self-explanatory.

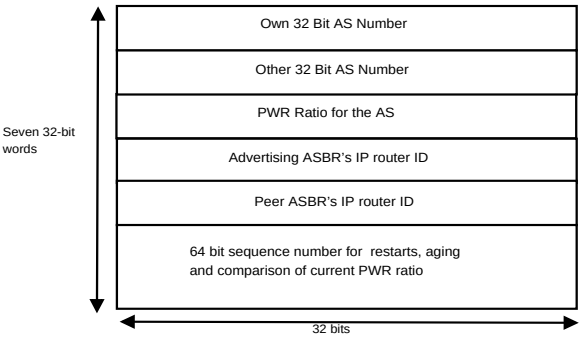


Fig. 9. AS-PATH-POWER-METRIC PDU