HotPLUZ: A BGP-aware Green Traffic Engineering Approach

Alejandro Ruiz-Rivera, Kwan-Wu Chin, and Raad Raad
University of Wollongong
NSW, Australia
Email: arr506@uowmail.edu.au, {kwanwu, raad} @uow.edu.au

Sieteng Soh
Department of Computing
Curtin University of Technology
Perth, Western Australia
Email: s.soh@curtin.edu.au

Abstract—Green networking techniques aim to shut down the least utilized links and/or routers during off-peaks hours. In this paper, we show that such techniques negatively impact the operation of the Border Gateway Protocol (BGP). We quantify the impacts of two representative green approaches: (i) GAES, a green technique that modifies link weights, and (ii) ESOL, a green technique that does not involve link weights adjustments. Experiments over the Abilene, AT&T, GEANT and SURFnet topologies show that when using GAES, routing changes and the proportion of rerouted traffic, both of which affect BGP, are in the order of 108% and 141% greater than ESOL. Therefore, we propose Hot Potato Low Utilization (HotPLUZ), a green approach that takes hot-potato routing into account. HotPLUZ reroutes traffic from lowly utilized links and aggregate said traffic onto highly utilized links, whilst minimizing any changes to the corresponding egress router of a given destination. In addition, HotPLUZ considers link utilization in order to avoid packet loss and high latencies. Our experimental results indicate an overall saving of up to 21% under low network load.

Index Terms—Green IGP-WO, green OSPF, hot-potato routing, BGP-aware, power savings.

I. INTRODUCTION

Green routing approaches aim to minimize the energy consumption of a network while preserving connectivity and guaranteeing network performance [1]. They typically involve shutting down lowly utilized links or routers during off-peak hours. However, to date, they have not considered the effect they have on the resulting cost of routes computed by an Interior Gateway Protocol (IGP). Figure 1 depicts two interconnected Autonomous Systems (ASes). The ingress router (IR), $R_1$, is presented with two equally good Border Gateway Protocol (BGP) [2] routes, via the egress routers (ERs) $R_2$ and $R_3$, with IGP cost $C_1$ and $C_2$ respectively. $R_1$ will choose the route with the lowest IGP cost, a policy known as early-exit or hot-potato routing. It is widely known that hot-potato routing causes numerous negative impacts on the operation of BGP. In [3] and [4], the authors establish three main impacts of IGP routing on BGP: (i) transient packet delay and loss while routers re-compute their forwarding tables, (ii) BGP routing changes affecting neighboring domains’ BGP operation, and (iii) shift in traffic that may cause congestion on new paths to other Autonomous Systems (ASes).

In this paper, we quantify the said impacts due to the use of green IGP Weight Optimizers (IGP-WO) [5] and green Open Short Path First (OSPF) [6]. Specifically, when using a green routing technique, we observe frequent route changes. In fact, our experiments indicate 29% more egress/border changes as compared to non-green approaches. Moreover, these techniques cause a significant, over 25%, traffic shift.

In light of the above results, we propose HotPLUZ. Its key features include: (i) minimizing the impact on BGP whilst minimizing overall energy consumption using the least cost IR-ER paths, ii) diverting traffic away from lowly utilized links and aggregate them onto highly utilized links, iii) consideration for link utilization to avoid packet loss and high latencies. Our simulation results over the Abilene, AT&T North America, GEANT and SURFnet topologies show that an overall saving of up to 21% is possible under low network load.

The rest of the paper is structured as follows. Section II presents related works. Section III introduces the implemented solution. Section IV describes the simulation methodology used to obtain the results presented in Section V. Section V quantifies the effects that green techniques may have on BGP and reports our experimental results for HotPLUZ. Finally, Section VI presents our conclusions.

II. BACKGROUND

A. Green IGP Approaches

We now review green techniques and highlight how they influence the IGP cost within a AS, and thereby, negatively impacting BGP’s operation. We start by discussing works that modify OSPF in order to reduce energy consumption. Even
though these techniques do not modify link weights per se, they do affect the performance of BGP as their modus operandi involves shutting down links.

Cianfrani et al. [7] introduce distributed Energy-Aware Routing (d-EAR). Their approach divides routers into ERs, IRs and Neutral Routers (NRs). Each ER is associated with a number of IRs. However, each IR is only associated with a single ER. Both ERs and NRs calculate a Shortest Path Tree (SPT). IRs on the other hand compute a Modified Path Tree (MPT) using their associated ER as the root node. The resulting MPT has fewer links and thus all links that do not appear in an IR’s MPT can be switched off. Consequently, IGP cost to BGP egress routers may be affected and the exit point for a given destination may change. Moreover, changes in IGP cost will force interior BGP (iBGP) to continuously recalculate routes. This leads to an increase in convergence time, packet loss and delay.

In a subsequent work, Cianfrani et al. [8] present an enhancement to d-EAR that considers positive and negative effects of exporting SPT. To this end, the term “move” is introduced to denote a set of strategies that place a target link into sleep mode. In particular, they consider a set of compatible “moves” that will minimize energy consumption. This process, however, may cause traffic that previously traversed current inactive links to be redirected, and thereby, changes the intra-domain traffic matrix. In turn, this may trigger BGP to select a different next hop for a given set of destination prefixes, causing a large number of BGP update messages. These message will overload routers’ CPU, cause transient packet loss and delay due to route re-computation.

Another green technique, called Energy Saving in the Internet based on Occurrence of Links (ESOL) [9], which runs over OSPF, quantifies the number of times that nodes and links appear in shortest paths using link state advertisements (LSAs). Links that are used less are marked as candidates to be powered off. After the removal of one or more links, ESOL will verify whether the network remains connected. In this regard, ESOL may shutdown links connected to BGP speakers. As a result, this will cause a sudden increase in traffic on other links, and have repercussions on outbound traffic and consequently the inbound traffic of a peer AS.

The following works modify the weight of links in order to maximize energy savings. The authors of [10] propose two Link Weight Optimizers (LWO) approaches, Greedy Algorithm for Energy Saving (GAES) and Two-stage Algorithm (TAES). In GAES, link weights are calculated initially using IGP-WO [11]. The objective function is the overall link utilization. Links are then sorted in non-decreasing order according to traffic intensity and link weight. The lowest ranked links become candidates to be switched off. GAES recalculates OSPF link weights without considering the links and nodes previously identified as candidates. The algorithm then determines if the given traffic matrix is supported without exceeding the links’ maximum utilization. In TAES, the algorithm first selects the set of network elements that could be switched off. This is carried out using an integer linear program (ILP) that minimizes power consumption. After that, the resulting topology is passed to IGP-WO, which then calculates new link weights that route current traffic demands subject to maximum link utilization constraints. These new link weights, however, may cause changes in intra-domain traffic, as was observed in [12]. Consequently, they may cause the selection of different egress routers for a number of prefixes.

Intra-domain traffic matrix changes are also likely to be caused by the approach presented by Arai et al. [13]. The authors propose ECO-Routing Protocol (ECO-RP). Here, a subset of routers are selected to gather and disseminate traffic volume information periodically using LSAs. Each router then modifies OSPF link weights dynamically according to traffic conditions. A subset of network elements, including nodes and links, is computed based on the new link weights. When network load is low, routers will route traffic to this subset of routers and place unused elements to sleep. On the other hand, when network traffic is high, routers readjust link weights so traffic can be routed over all nodes and links. As discussed in [14] [15], the continuous change in link weights may generate a large number of BGP update messages, which may overload routers. Moreover, a sudden traffic shift is also expected.

Changes in traffic conditions are also explored in [16] with the introduction of Routing On Demand (ROD). Using non-linear optimization techniques, the researchers combine conventional traffic engineering (TE) and green approaches in order to maximize energy savings and network performance under the constraint of maximum link utilization (MLU). The aim is to minimize the risk of network congestion. In particular, their green TE approach determines the minimum set of links capable of routing a given traffic demand. The algorithm, therefore, combines both conventional and green TE techniques evaluated under the same multiple commodity flows (MCF) constraints. By introducing a “green factor”, ROD relies on a centralized controller that periodically computes link weights and configure routers. In order to direct traffic to specific links, this approach requires continuous route re-computation, which may cause transient packet delay and packet loss as argued in [3] and [14]. Hence, continuous changes in traffic demands are likely to cause adverse BGP impacts.

In summary, we can divide the previous works into two categories: i) **OSPF changes**, such as d-EAR [7], [8] and ESOL [9]. These techniques are likely to cause an increase in BGP update messages, leading to high CPU load on routers and frequent re-computation of routes along with a potential rise in the amount of inter AS traffic, and ii) **LWO approaches**, like GAES, TAES [10], ECO-RP [13] and ROD [16]. These approaches induce changes in intra-domain traffic, which may cause the selection of a different egress router for destination prefixes. In addition, they may cause more BGP update messages, and the resulting LSA flooding may overload routers. Transit packet delays and packet loss are also expected due to continuous changes in traffic demands.
B. Minimizing BGP Changes

We now briefly explore the most relevant works that consider the implications of BGP changes. Routing changes affecting BGP next hop selection are explored in [15]. Balon et al. argue that changes in IGP cost towards a particular BGP speaker may lead to traffic shifts. When a traffic shift indeed takes place, this results in a change to the intra-domain traffic matrix. In order to reduce the effect that IGP cost changes have on the intra-domain matrix, a BGP-aware Link Weight Optimizer (BGP-LWO) [14] [12] is developed using the objective function presented in [11]. This objective function assigns a large value to highly loaded links with the aim of minimizing the total link load of the network.

BGP-LWOs like the one explored in [15] are however designed to be deployed in iBGP full-mesh configurations, where each route announcement received by a router on a BGP session is retransmitted on iBGP sessions. In this scenario, every router is said to have Complete Visibility (CV) of all the available routes for every destination and it can choose its best global route [5].

A BGP-LWO that requires CV is presented in [12]. The paper notes that by modifying link weights in order to minimize a network-wide objective function for a given intra-domain traffic, BGP may be affected, as the resulting set of link weights is likely to trigger BGP to change the BGP next hop for a given set of destination prefixes, causing degradation in network performance. Changes in link weights can also lead to an increase in load on some links as a result of traffic shifts.

The reviewed works thus far indicate that IGP LWO affects the final intra-domain traffic matrix. Awarwal et al. [14], however, go further by quantifying and measuring the resulting traffic due to link weight optimization. The authors conclude that techniques that ignore hot-potato effects can result in IGP metrics that are suboptimal by 20% in regards to link utilization. They also report as much as 25% shift in traffic to a remote AS due to link weight recalculations.

III. Hot Potato Low Utilization (HotPLUZ)

From the aforementioned works we can establish a list of design considerations that a green technique should take into account in order to avoid negatively impacting BGP. Note, these considerations are similar conceptually to those put forth by Teixeira et al. [17], which only consider minimizing hot potato changes as opposed to energy consumption:

1) IRs must select their optimum egress points. In our case, optimum refers to the shortest IR-ER path, which accomplishes the objective of hot-potato routing.
2) IRs use only one path per ERs. This aims to reduce the size of BGP routing tables and the number of BGP update messages. The consideration also aims to reduce re-routing computation and hence the routers CPU processing.
3) Tunnels are established between IRs and their chosen ER. The purpose of these tunnels is to ensure IR-ER path stability, which reduces IGP changes and increases BGP stability.

As mentioned in Section I, the main goal of green routing approaches is to conserve power by switching off routers and links. However, this needs to be carried out without negatively impacting BGP. Given the design considerations, we propose a distributed approach that allows a network operator to switch off links to conserve power usage. In addition, the approach ensures IRs are able to establish a connection to their respective ERs and the maximum link load is minimized. The latter is important as a high load will lead to increased delays and possibly packet loss, despite great power savings.

Before outlining HotPLUZ, we first define a few terms and key concepts. We will use $V$ to denote the set of nodes connected by edges in the set $E$. Each link $e \in E$ has capacity $c_e$ and $L'_e$ is defined as the sum of all traffic traversing such link at any given moment, i.e., link utilization. An IR-ER pair is represented as an $(r, s)$ pair with a demand of $d_{rs}$. Let $P_{rs}$ be the set of all simple paths for pair $(r, s)$. Note, an IR will have a different set of paths for each ER. For each IR $r$, we will denote its chosen path for ER $s$ as $p^*_{rs}$. We now define how link cost, $w_c$, is calculated. Initially, every link $e \in E$ has an IGP cost of $1/c_e$. Subsequent IGP costs are recomputed based on link utilization using Equ (1), where $(c_e - L'_e)$ represents the available bandwidth of a given link $e$. The use of Equ (1) causes IRs to divert traffic from lowly utilized links and aggregate said traffic onto highly utilized links. This is carried out with the goal of switching off lowly utilized links.

$$w_e = \begin{cases} \frac{1}{1-(c_e-L'_e)\gamma}, & 0 < L'_e \leq MLU \\ \infty, & L'_e > MLU \end{cases}$$ (1)

$$ALU = \sum_{e=1}^{\left|E\right|} \frac{L'_e}{\left|E\right|}$$ (2)

To avoid congestion and packet loss on links, we define a maximum average link utilization (MALU) threshold, where the average link utilization (ALU) is calculated using Equ (2). HotPLUZ also restricts link utilization, $L'_e$, to a predefined maximum link utilization (MLU) value; i.e., $L'_e \leq MLU$. This MLU value is a percentage of the total capacity of a given link $c_e$; i.e., $\gamma c_e$, with $0 \leq \gamma \leq 1$. These two parameters, MALU and MLU, are predefined by the network operator. Also, the MLU can be set to a value that allows a link to absorb any sudden burst in traffic. In particular, as per [18], to keep delay low, the MLU is set to 80% and the MALU to 70% as per [19].

We are now ready to describe HotPLUZ. It works in rounds, and consists of the following key ideas:

- In round zero, all IRs select their corresponding ERs using hot potato routing, whereby they select a BGP speaker or ER whose intra-domain distance or IGP cost is the smallest.
- In subsequent rounds, each IR determines the link cost as per Equ (1). After that, for each IR, it determines whether there exists a new least cost path to a corresponding ER.
- If there is, the IR changes to the new path.

The above process continues until no IR makes any changes. That is, all IRs converge to the least cost path for each
corresponding ER. Upon completion, the network may have a number of links with zero utilization. The network operator then has the option to switch these links off.

Figure 2 presents the flow chart of HotPLUZ. Initially, HotPLUZ computes the shortest IR-ER paths based on hop-count. Every IR then sends an UPDATE message to other IRs to inform them its path selection and corresponding demands. Once an IR receives an UPDATE message, it uses said information to calculate new link weights as per Equ. (1) and computes possible new least cost paths. If one is found, the IR then establishes the new least cost path, \( p^*_rs \). An UPDATE message is then sent to other IRs and the process repeats until any of the following three conditions occur: i) The maximum predefined link utilization, MALU, is reached, ii) none of the paths \( p \in Prs \) are able to carry the given demand, or iii) when there are no more changes in the selected (\( r, s \)) path.

![Flow chart for HotPLUZ](image)

**Fig. 2: Flow chart for HotPLUZ**

### IV. Simulation Methodology

This section presents the simulation methodology used to obtain our results. It is divided into two parts. The first part explains how we calculate the proportion of egress router changes and the resulting percentage of re-routed traffic within a single AS. The second part presents how we validate the performance of HotPLUZ.

#### A. Quantifying Hot-Potato Routing Effects

We study two green approaches: i) Greedy Algorithm for Energy Saving (GAES) [10]. This technique shuts down the least utilized links while taking into account the average network load whilst maintaining network connectivity. In our experiments, initial link weights are calculated using IGP-WO [11] – available as part of the TOTEM toolbox [15]. The objective function of IGP-WO is the overall link utilization, which is directly related to network congestion. After the initial link weight calculation, links are sorted in non-decreasing order according to their utilization and the least loaded links are switched off. The algorithm then determines if the given traffic matrix is supported by the remaining links; i.e., no links exceed their maximum utilization, and ii) Energy Saving in the Internet based on Occurrence of Links (ESOL) [9]. This technique represents green approaches that only place links to sleep without modifying link weights. ESOL runs on top of OSPF and uses LSAs to determine the occurrence of nodes and links in all calculated shortest paths. Network links with low frequency are placed into sleep mode. ESOL also considers network connectivity and maximum link utilization when powering down links.

We conducted experiments over the following topologies: SURFnet (50 nodes, 138 links), GEANT (40 nodes, 122 links), AT&T (25 nodes, 112 links), Abilene (11 nodes, 28 links) [20] [21] [22]. These networks are constructed in TOTEM [15]. In order to simulate a Multi-Protocol Label Switching (MPLS) network, each bidirectional connection between nodes is replaced by unidirectional links with similar capacity. This guarantees that each link is assigned the same TE link weight. We assume that our MPLS network runs OSPF. This is to allow the deployment of TE-based green techniques [23]. We then generate a traffic matrix (TM) using the well known gravity model [24] [25].

Once a given topology and its respective TM have been simulated in TOTEM, we run GAES and ESOL. In the case of GAES, we run IGP-WO. This algorithm calculates the optimal link weights in order to route the given TM. The output is a network that contains fewer links capable of supporting the given TM. In the case of ESOL, OSPF is run without performing any modifications to link weights. Similar to GAES, it generates a topology with fewer links that can support the given TM. In order to minimize the number of changes in IGP costs, both approaches on each iteration only eliminate five links; the least loaded for GAES, and the ones with low occurrence in shortest paths for ESOL, with the constraints that the overall link utilization is below 50% [26] [27] and the resulting network remains connected. At the end of all the iterations, GAES and ESOL generate as output a final topology. We then record the remaining links, and link weights in the case of GAES. We then run the following steps 100 times:

1. We select a random number of routers to be part of the following two groups: i) Egress routers, which act as exit points, and ii) Ingress routers, which use the said egress routers in order to reach a particular destination outside the local AS.
2. Using the original topologies, i.e., ones which have not been processed by GAES or ESOL, we run Dijkstra’s algorithm, and calculate for each ingress router the distance to each egress router. The egress router with the minimum distance is then selected as the exit point for that specific ingress router. We also record the traffic being sent to the selected exit point.
3. Lastly, we repeat Step 2 using the GAES and ESOL...
topologies.

4) We compare the results obtained in Steps 2 and 3 and calculate the portion of egress routers that have changed for a particular set of ingress routers. We also record the proportion of traffic that needs to be re-routed.

B. Evaluation of HotPLUZ

In addition to the aforementioned four topologies, namely SURFnet, GEANT, AT&T and Abilene, we constructed a topology called Random_8 (8 nodes, 24 links), which we generate using the TopGen built-in function of TOTEM [15]. Traffic matrices (TMs) are again obtained using the gravity model. For each topology, we first generate a base TM, called $TM_{Base}$, and then another two TMs are calculated by multiplying $TM_{Base}$ by 5, and 20. Each node is then randomly classified as an ER or IR. After that each IR selects by multiplying $T M$ called model. For each topology, we first generate a base TM, we conducted 50 simulation runs for different number of ERs, i.e., $num$, $num = 2, 4, 6$. In all cases, an IR selects new path as long as the new path has sufficient bandwidth to support a given IR-ER traffic demand. We only consider IR-ER traffic. This means links that do not carry IR-ER traffic have zero utilization.

The savings due to HotPLUZ is defined as $(Z - A) / |E| * 100$. The overall saving for each topology is calculated by averaging the final energy savings for each TM; i.e., $TM = 5 * TM_{Base}, 20 * TM_{Base}$. In addition, we report the average utilization achieved by HotPLUZ and the average number of iterations that HotPLUZ needs in order to reach stability.

V. Results

We will first present results that quantify the negative impact of existing green techniques before outlining the performance achieved by HotPLUZ.

A. Hot-Potato Routing Effects

Figures 3a and 3b show the percentage of changes in a path or cost to an existing egress router, aka, “hot-potato routing change”, for GAES and ESOL respectively. Our results indicate that the proportion of re-routed traffic is directly related to the percentage of changes in selected egress routers. This percentage along with the proportion of traffic that needs to be re-routed depends on the number of links that are switched off and continuous changes in link weights, causing IGP cost recalculation leading to long convergence delays.

Figures 3a and 3b show that for all simulated topologies, the percentage of hot-potato routing changes for GAES is 32.5% of ingress routers select a new egress router after the implementation of GAES and ESOL respectively, which causes a corresponding 48.4% and 24.7% shift in traffic. This large shift in traffic is likely to cause undesired effects such as external BGP routing changes [17].

Our results, figures omitted due to space limitation, show the proportion of re-routed traffic for GAES is greater, i.e., 141%, than for ESOL. This discrepancy can be explained by the fact that ESOL does not modify link weights, and only shuts down links. Therefore, traffic does not get re-routed as much as in GAES. That is, ESOL causes fewer changes in IGP cost and hence shortest paths between nodes do not change often. In addition, we observed that Abilene has the highest proportion of re-routed traffic, with 25.3%. Abilene is the smallest among the four simulated topologies and therefore, the number of link changes in computed shortest paths due to the shut down of links are more severe than in other topologies.
the algorithm has to run, i.e., more paths to choose from, the more energy reduction is achieved. For large topologies with rich connectivity, at low network load, HotPLUZ has more opportunities to reduce energy consumption but it will also need more iterations to stabilize. Recall that our HotPLUZ adheres to link utilization constraints, which help prevent excessive aggregation of traffic on certain links which can cause packet loss and high latencies.

VI. CONCLUSIONS

We have shown how green networking approaches have a non-negligible impact on hot potato routing, and thus are likely to negatively interfere with BGP’s operation. That is, current green IGP techniques run the risk of exacerbating BGP convergence delays, and degrade the QoS of peer AEs. This conclusion is drawn from our study of two representative green IGP-WO and OSPF techniques, namely GAES and ESOL. We have quantified their impact on the proportion of egress router selection and percentage of re-routed traffic within a single AS and routes advertised by peer AEs. Our results show that for GAES, hot-potato routing changes and the proportion of re-routed traffic are in the order of 108% and 141% greater than ESOL, respectively. These results thus serve as motivation for HotPLUZ, a distributed approach that switches off links without negatively impacting BGP. Our experiments show that up to 21% overall saving are achieved under low network load. As a future work we are interested in exploring collaborative solutions between different AEs running BGP-aware green network approaches.

REFERENCES


TABLE I: Overall saving (%) achieved by HotPLUZ for values of num = 2, 4, 6 and under different network loads.

<table>
<thead>
<tr>
<th>TM</th>
<th>Random_8</th>
<th>Abilene</th>
<th>AT&amp;T</th>
<th>GEANT</th>
<th>SURFnet</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 * TM_Base</td>
<td>3.7</td>
<td>6.8</td>
<td>21.7</td>
<td>11.5</td>
<td>9.6</td>
</tr>
<tr>
<td>20 * TM_Base</td>
<td>3.7</td>
<td>0</td>
<td>21</td>
<td>6.9</td>
<td>1.8</td>
</tr>
</tbody>
</table>

TABLE II: Average link utilization (%) exhibited by HotPLUZ for values of num = 2, 4, 6 and under different network loads.

<table>
<thead>
<tr>
<th>TM</th>
<th>Random_8</th>
<th>Abilene</th>
<th>AT&amp;T</th>
<th>GEANT</th>
<th>SURFnet</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 * TM_Base</td>
<td>1.0</td>
<td>10.3</td>
<td>1.9</td>
<td>1.5</td>
<td>2.8</td>
</tr>
<tr>
<td>20 * TM_Base</td>
<td>3.7</td>
<td>40.1</td>
<td>9.4</td>
<td>6.0</td>
<td>9.8</td>
</tr>
</tbody>
</table>

TABLE III: Overall number of iterations for values of num = 2, 4, 6 under TM = 5 * TM_Base and TM = 20 * TM_Base

<table>
<thead>
<tr>
<th>TM</th>
<th>Random_8</th>
<th>Abilene</th>
<th>AT&amp;T</th>
<th>GEANT</th>
<th>SURFnet</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 * TM_Base</td>
<td>0.3</td>
<td>0.78</td>
<td>2.5</td>
<td>1.5</td>
<td>1.7</td>
</tr>
<tr>
<td>20 * TM_Base</td>
<td>0.3</td>
<td>0</td>
<td>2.5</td>
<td>0.7</td>
<td>0.2</td>
</tr>
</tbody>
</table>