Power-proportional Router: Architectural Design and Experimental Evaluation

Bin Liu, Jianyuan Lu, Yi Kai, Yi Wang and Tian Pan
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
{lmyujie, thlujy, kaiyi02, pig020623, platinum127}@gmail.com

***Abstract*—High speed routers in Internet are becoming increasingly more powerful, as well as more energy hungry. However, they always show power-inefficient property due to we unilaterally in pursuit of high speed before. In response to this problem, we present a power-efficient router architecture named GreenRouter in this paper. GreenRouter separates a line card into two parts physically: the network interface card (named as DB) and the packet processing card (named as MB), which are interconnected by a two-stage unidirectional switch fabric. Traffic from all the DBs shares all the MBs in GreenRouter, thus the traffic can be aggregated to a few active MBs when traffic is light and the inactive MBs can be shut down to save power. We give the detailed architectural design of GreenRouter. Real-trace driven experiments show that GreenRouter can save about 50% power compared to the conventional router when the average traffic load is 30%, while providing quality of service guarantee at the same time.**

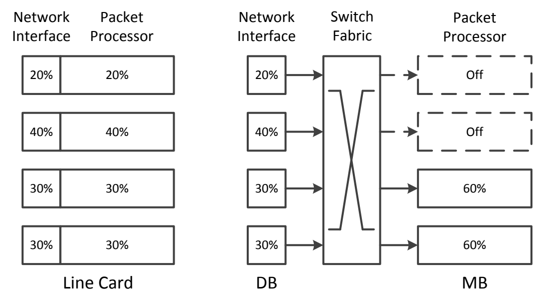
I. Introduction

Recent studies reveal that the Internet transmission links exhibit relatively low average utilization. The maximal average utilization of today's backbone links is reported to be 40% or even less [1]. And network traffic normally has the 24-hour-effect, higher in the daytime and lower at night [2, 3]. But routers are conventionally designed to meet the worst case traffic demand which results in low power efficiency.

In this paper, we propose to redesign a power-proportional router at the architectural level named *GreenRouter*. Our vision is to let all the interfaces share all the processing engines/capacities dynamically and adaptively create room for more power saving. An intuitive example is shown in Figure 1. Generally, a line card in a conventional high-end router consists of two major parts:

- *Mother Board (MB)* for packet processing, including IP longest prefix match, traffic classification, packet/flow queuing and scheduling and so on.
- *Daughter Board (DB)* consisting of network interfaces.

Both parts are tightly bundled together, either physically (in the same printed circuit board (PCB)) or logically (in separate PCBs but grouping into one whole entity). Therefore, from the view point of architectural design, the conventional routers cannot effectively take advantage of the network traffic conditions due to their poor power-proportionality. Just as shown in Figure 1(a), we assume a four-port conventional router, all the line cards should keep working (power-on) although their utilizations are low.



(a) Conventional Router (b) GreenRouter

Fig. 1. An Illustration of Power-proportionality Comparison

By contrast, in GreenRouter, the DBs and MBs are physically separated and interconnected via a switch fabric, all the MBs can be viewed as a resource pool which can be shared by all the DBs. In this way, the traffic from a DB can be split to distribute among all the MBs on-demand. In the example of Figure 1(b), the traffic from the external interfaces are aggregated into only two MBs, the other two MBs can be powered-off to save power. In this case, each active MB carries 60% traffic load, so the quality of service (QoS) can still be well guaranteed. As mentioned in [4], MBs are the dominant power consuming part in a high-end router, so we can potentially achieve larger power saving benefit than that in traditional routers.

Although the idea of sharing processing capacities appears simple, we will meet several research challenges in designing GreenRouter:

1) How to interconnect MBs and DBs in order to effectively share MBs' processing capacities?
2) How to dispatch traffic from DBs to active MBs?
3) When to sleep or wake up an MB with traffic fluctuation?
4) How to guarantee GreenRouter's performance under pre-defined QoS constrains?
5) How to maximize the power saving percentage in GreenRouter?

To address these challenging issues, we provide both architectural design and experimental evaluation by real trace driven simulations.

The rest of the paper is organized as follows. We compre-

This work is supported by 863 project (2013AA013502), NSFC (61373143), Tsinghua University Initiative Scientific Research Program (20121080068). Corresponding Author: Bin Liu (lmyujie@gmail.com).

hensively survey the related work in Section II. In Section III, we present the architectural design of GreenRouter. In Section IV, we evaluate GreenRouter's performance including power saving benefit via real trace simulations and in Section V, we conclude the paper.

II. Related Work

The survey of previous studies is divided into two directions: (1) Power saving in traditional routers and; (2) Power saving in architectural redesigned routers. Despite of the dividing, most of the power saving approaches in the former can be applied to the latter.

A. Power saving approaches on existing routers

In the context of saving existing routers' power consumption, the literature can be roughly classified into three categories: 1) Sleeping/standby network element, 2) Dynamic rate adaptation and, 3) Dynamic frequency scaling (DFS) and dynamic voltage scaling (DVS).

1) Sleeping/standby network element at node/link level: Sleeping and standby approaches are used to drive unused network elements to low energy modes. The behind idea lies in that a transmission link is not always busy, sometimes it is idle when being lightweight. In this case, when a link is in low traffic, we can buffer the arrival packets for a while to sleep the processing engine. According to this idea, there are two methods along this direction: a) Uncoordinated manner and, b) Coordinated manner. For the former, the representative study is the B&B method [5]. In [5], Nedevschi et al. proposed the buffer-and-burst approach which shapes traffic into small bursts to create greater opportunities for network components to sleep. For the latter, the representative work is [6]. In that paper, Mingui et al. proposed GreenTE to achieve power-aware routing. They idle some links to go sleep to save power. However, it needs network wide coordination and currently is mainly limited to within an ISP.

2) Dynamic rate adaptation at link level: Dynamic rate adaptation approaches are designed to modulate capacities of network resources in order to meet the current ties and link bandwidth. This approach is suitable only for Ethernet links. The current energy efficient Ethernet (EEE) research basically belongs to this category [5, 7–12]. Nedevschi et al. [5] showed that even simple schemes for rate adaptation can offer substantial power savings without noticeably increasing loss and with a small increase in latency. However, most Internet backbone transmission links use SDH/SONET specification and link-rate adaptation does not apply any more.

3) Dynamic frequency scaling (DFS) and dynamic voltage scaling (DVS) at chip level: DFS and DVS are the two commonly used techniques in saving power at chip level when adapting the arrival load [13–16]. Some more recent work has also been done by Kennedy et al. [17, 18] and Meng et al. [19]. But we can only get small energy saving benefit from these two approaches. The reason lies in two aspects: (i) In the modern routers/switches, the chips have already used very low voltage such as 0.9V; (ii) The changes in frequency do not provide as dramatic power saving as change in voltage.

B. Power-efficient router design via architectural innovations

In terms of architectural redesign, in a short paper [20], Yamada et. al. proposed a centralized architecture router with multiple shared packet processing engines. But it does not touch issues like the actual architectural design, packet scheduling, and fast processing card wakeup. Bianco et al. [21] described a distributed software router that has a front-end interface server and multiple back-end packet processing servers. It improves energy proportionality in a fashion similar to GreenRouter but only for software routers. In [21], it said nothing about avoiding the packet out-of-order issue, which is a major concern when sharing the servers by all the clients. To the best of our knowledge, we are the first to study architectural redesign for power-efficient high-end routers by unbinding MB and DB in a line card and sharing the MBs' processing capacities explicitly as a resource pool. A fast letter [22] published the basic idea but has not given out detailed architecture and the experimental evaluation is simple.

In other related aspects. The two-stage load-balanced switch was proposed by [23]. However it fixed the first switch to be N×N and this is very different from our traffic distributing case. Ours is that the number of destination servers (MBs) is changeable on-demand as traffic fluctuations. Similarly, the well-studied traffic load-balancer such as [24] targeted at only evenly distributing the incoming packets to a fixed number of servers, regardless of the dynamic changed server number.

III. Power-Proportional Architecture of GreenRouter

A. Architecture of Conventional Routers

Figure 2(a) shows the interconnection structure of conventional high-end routers. Such a router mainly consists of three kinds of cards: 1) Route Processing Card. This card is in charge of executing routing protocols and producing routing tables. Besides, it also manages the whole system resources; 2) Line Card. This card is responsible for terminating packets from external link(s) and forwarding packets upon searching lookup table in the data plane; 3) Switch Fabric Card. Usually an N×N switch fabric is built in a router to interconnect line cards. A router also contains other auxiliary modules such as fan system for cooling and power supply system.

TABLE I
Breakdown of Cisco router's power consumption (from [25])

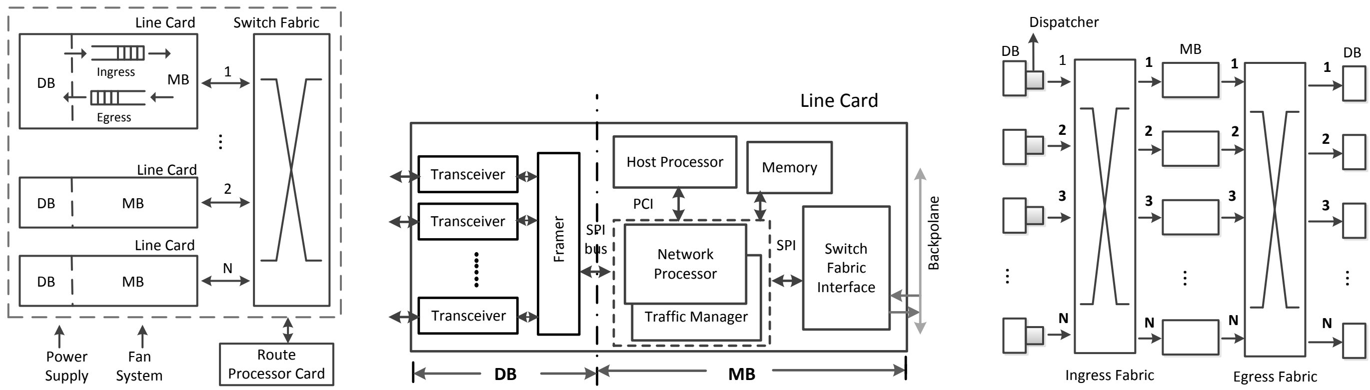
(a) Router

router component	idle	peak
line cards	78%	67%
fan trays	7%	18%
switch fabric	3%	3%
power suppiles	9%	9%
route processor	3%	3%

(b) Line card

line card component	power
ASICs/NPs	53%
local power supplies	14%
CPU/DRAM	9%
interconncet	7%
others	4%

In conventional routers, each line card is a functionally indiscerptible binding entity and includes two major functional parts: i) Interface part. It mainly consists of transceiver in the



(a) Interconnection Structure in a Conventional High-end Router

(b) Composition of a Line Card

(c) GreenRouter Interconnection Structure

Fig. 2. Architectural Comparison between the Conventional Router and GreenRouter

physical layer (like PHY in Ethernet spec) and the framer in the data link layer (like MAC chip in Ethernet spec). This part is basically corresponding to the DB in GreenRouter; ii) Processing part. This part mainly consists of: a) network processor for packet parsing, route lookup/classification and manipulations; b) traffic manager for packet/flow queueing/scheduling; c) host processor, which is an embedded onboard CPU system responsible for communicating with route processor and managing the local hardware; d) Front-end switch fabric interface, which is responsible for connecting this line card to switch card(s) via backplane. This part is basically corresponding to the MB in GreenRouter. These two parts are connected via standard SPI-bus. Please note, a line card runs the duplex processing, i.e. ingress and egress processing.

We have measured a commercial router, Huawei's NE40E-X8, to examine the decomposition of its power consumption. Under a configuration of 8 line cards, 1 switching fabric, and 2 route processor cards (a master card and a backup card). We got: (i) the line cards account for 68% the total system power consumption; (ii) the idle power when no traffic is more than 97% of its peak power(fully loaded); (iii) the interface module only consumes 6.8% of a line card's whole power consumption. The similar results were also reported by Wobker from Cisco [25], see Table I. So we can conclude the following:

- Line cards are the dominant power consumer in a router.
- A line card's idle power is close to its peak power.
- A line card's interface port only consumes a small amount of power.
- A switching fabric only consumes a small amount of power.
- Fan speed, which is automatically regulated based on internal temperature, can affect power consumption.

B. Resource Pool Sharing Mechanism and Power-proportionality in GreenRouter

The core idea of GreenRouter is to split a line card into an interface part (DB cards) and a processing part (MB cards), keep all the DBs up and running, but make the MBs a service pool and manage them according to the workload. If a router's processing capability can adapt to traffic level, other parts (such as fans) will be able to adapt too, which eventually will lead to elastic, energy-proportional routers. To this end, we redesign GreenRouter's architecture shown in Figure 2(c).

In Figure 2(c), we use a two-stage unidirectional switch fabric to interconnect DBs and MBs. The first stage switch fabric (also called ingress fabric) actually acts as a traffic distribution network, while the second stage switch fabric (also called egress fabric) works as a routing network. These two fabrics exactly share the same techniques with the existing switch fabric design in conventional routers. We add on a new functional unit, named *Dispatcher*, in DB's ingress direction to distribute traffic from interface to multiple active MBs under the constraints of both traffic load-balancing and minimizing the number of active MBs (meaning maximizing the power-saving. The dispatching policy will be discussed in the following section). The functionality of DB and MB is basically the same partition as indicated in Figure 2(b). Unlike the processing resource provisioning in conventional routers, where the packet processing hardware in a line card is dedicated to its own interface(s), in GreenRouter, all MBs are fully shared by all DBs, making MBs a resource pool structure.

When a packet comes from an external interface of GreenRouter, it will first go through L1/L2 processing circuits to recover a complete IP-formatted packet, then be encapsulated into the internal transmission format between DB and MB going through the first stage switch fabric, and then arrives at a certain MB. After being processed by the MB, this packet will be forwarded to its final destination port in DB's egress direction via the second switch fabric.

In GreenRouter, we introduce three new components: (i) a dispatcher in each DB; (ii) an additional switch fabric; and (iii) a control module. The logical connection among these components is shown in Figure 3. Control module is a small software patch and designed to reside in route processor card.

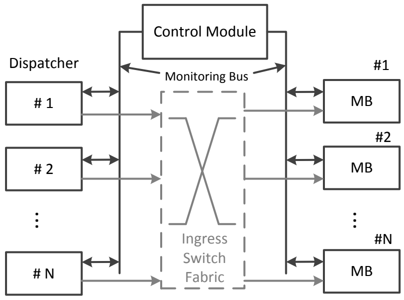


Fig. 3. Connection among Dispatchers, MBs and Control Module

It collects the MBs' information such as resource utilization via the monitoring bus and makes decision to sleep/wake up an MB. It also notifies the DBs about the current status of MBs in the similar manner. When the link load is light, the dispatcher will aggregate/distribute incoming traffic into fewer active MBs and correspondingly, the control module will idle other MBs to sleep for power-saving. Contrarily, when the link load goes higher, the control module will wake up more MBs and dispatchers will be noted about this information for adjusting traffic distribution. In this way, GreenRouter will automatically adapt the external traffic fluctuation, thus well support the elastic flexibility in resource provisioning to achieve good power proportionality.

GreenRouter will keep all DBs power-on and thus maintain a non-interrupted communication with outside network. In this sense, GreenRouter is fully transparent to the existing Internet and asking no change on the current network protocols.

IV. Experiment and Evaluation

In this section, we evaluate GreenRouter's power saving gain, packet reordering as well as the packet delay performance by applying real traffic traces. The results show that it is able to achieve considerable power savings with little impact on QoS performance. Due to space limitations, the traffic dispatching algorithm and modeling analysis are omitted. Details can be found in a technical report [26].

We simulate the functionality of GreenRouter, in order to check if GreenRouter's functionality works properly as we expect. In the simulations, traces from different ports arrive at GreenRouter at the same time, and the distributed dispatchers schedule the traffic in parallel. Packets are dispatched and processed according to the real-time status of the system.

A. Evaluation Metrics

The following metrics are tested during these simulations:

- *Amount of active MBs*. We need to examine if the number of active MBs adapts to the fluctuations of incoming traffic. Besides, the active MB number is used to estimate power saving effect.
- *Reordering ratio*. When a packet belong to an existing flowlet arrives and the corresponding MB is full, we have to choose another MB to process this packet. In this situation, reordering of a flow may take place.
- *Extra delay*. An extra buffer is used at the entry of an MB. This might bring extra delay comparing to conventional routers. We monitor every packet's extra delay directly in simulations.

TABLE II
Information and Power Saving of Original Traces

Trace ID	Collect Point	Average Load	Power Saving
1	Chicago	31.61%	59.30%
2	San Jose	78.07%	6.01%
3	CERNET	52.68%	33.73%

TABLE III
Information and Power Saving of Modified Traces

Trace ID	Collect Point	Average Load	Power Saving
1	Chicago	19.27%	63.70%
2	San Jose	53.02%	31.22%
3	CERNET	31.60%	51.70%

B. Experimental Settings

Three different real traces for the evaluations are used as shown in Table II. All the three traces last 60 minutes. Two of them are collected from backbone links of a Tier1 ISP from CAIDA [27, 28], another is collected from the international egress link of China Education and Research Network (CERNET).

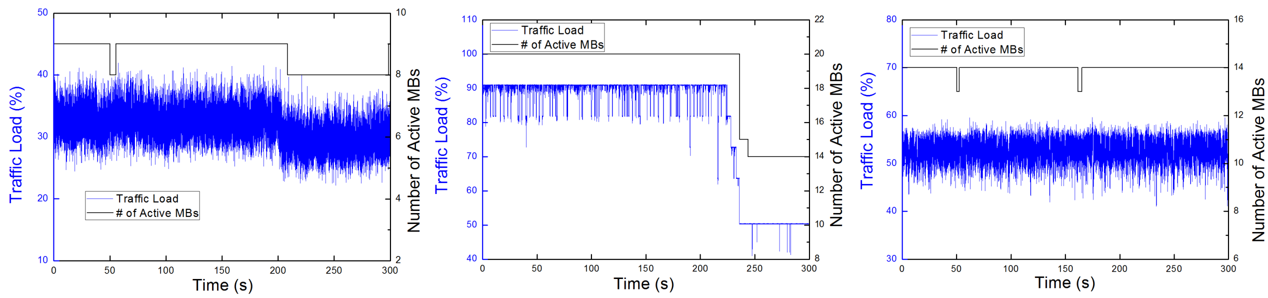
We are unable to get traces from a router's all ports. But we want to simulate a multi-port router. So every trace is cut into 20 pieces and a 20-port router is constructed. These 20 pieces are assumed to arrive at GreenRouter's 20 different ports at the same time.

On the other hand, lengths of the traces which we can get are not long enough. It is hard to evaluate the power saving effect of GreenRouter in a long period. The impact of 24-hour-effect, which means traffic volume may be heavy at daytime but very light at night, is very common on both domestic and international links [2], and it is a great opportunity for GreenRouter to save power. We modify the throughputs of these three original traces to imitate the 24-hour-effect. Sine wave filter for throughput is used to modify traces. Although sine wave does not fit the 24-hour-effect very accurately, adaptation of active MBs can be observed. The modified traces' information is shown in Table III.

C. Simulation Results

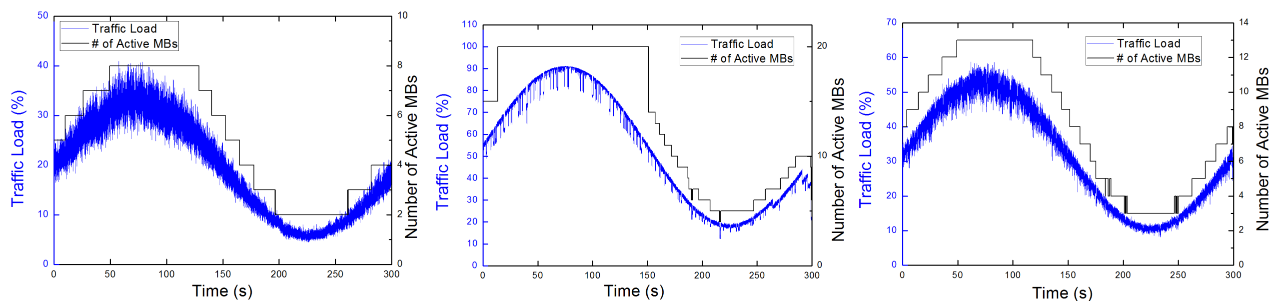
1) Power Savings

Figure 4 illustrates the results of experiments using three original traces respectively, where the number of active MBs varies while traffic load changes although the fluctuation is not obvious. Compared with conventional routers, GreenRouter expenses less power, due to parts of the processing capacity being powered-off. The number of active MBs is used to estimate GreenRouter's power consumption. Table II summarizes power saving using three original traces. When the average utilization is low, as trace 1 and 3, GreenRouter can save



(a) Chicago (b) San Jose (c) CERNET

Fig. 4. Simulation Results: Dynamic Number of Active MBs with Original Traces.



(a) Chicago (b) San Jose (c) CERNET

Fig. 5. Simulation Results: Dynamic Number of Active MBs with Modified Traces.

TABLE IV
COMPARISON OF POWER SAVING BETWEEN FLOWLET-BASED AND PACKET-BASED RESULTS

Trace	Average Load	Packet-based	Flowlet-based
Chicago	19.27%	64.31%	63.70%

significant power. But when the average utilization is high as trace 2, the power saving gain is accordingly decreased.

Figure 5 illustrates the results of experiments using three modified traces. And Table III summarizes power saving gain using these three modified traces. Adaptation of active MBs is more clear in Figure 5. We can also learn from Figure 5 that processing capacity of GreenRouter is always a little higher than the traffic volume, because of the reserved processing capacities. This is a tradeoff between QoS performance and power saving gain.

Packet-based dispatching can minimize the power consumption of GreenRouter, without considering QoS constraints. Comparative experiments are conducted between flowlet-based and packet-based dispatching (definitions can be found in [26]). We select the trace from Chicago to implement the experiments. Figure 6 and Table IV illustrate the experiment results. Flowlet-based dispatching is less power efficient than packet-based dispatching, but they are very close. That means power consumption of GreenRouter with flowlet-based dispatching policy is very close to the best power saving gain, while QoS constraints are considered.

2) Reordering Ratio and Extra Delay

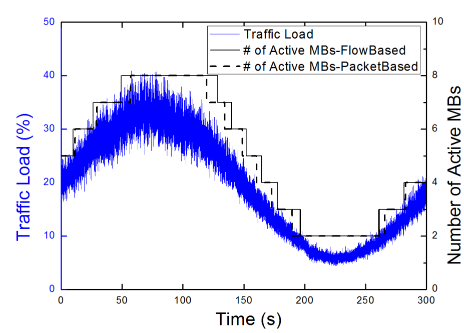


Fig. 6. Comparative between Flowlet-based and Packet-based

The size of the buffer at an MB's entry is a key factor of reordering ratio. We do the experiments using the three original traces with different buffer size. Figure 7 illustrates the reordering ratios. The reordering ratio decreases rapidly when buffer size increases. It can be ensured to be small enough, e.g. 10^{-8}, by setting appropriate buffer size. While applying buffer to handle traffic fluctuations to decrease the flow reordering ratio, extra delay is also introduced. Generally, larger the buffer size, smaller the flow reordering ratio but longer the extra packet delay. When designing GreenRouter, we make the trade-off between these two. The maximal extra packet delay is determined by both the buffer size and the MB's processing time.

Figure 8 illustrates the CDFs (Cumulative Distribution Function) of the extra packet delay over the three original traces. The buffer size is set to be 100 packets, each packet

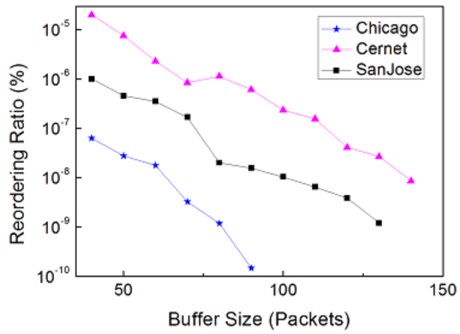


Fig. 7. Reordering Ratio

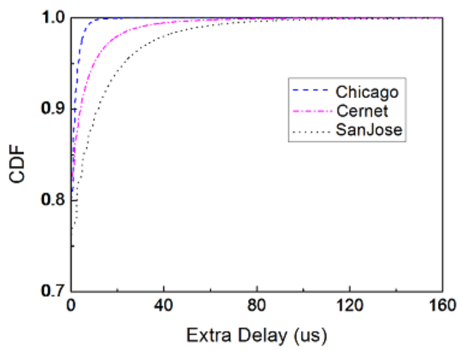


Fig. 8. CDF of Extra Delays

1500 bytes. Most of the packets experience very low extra delays, and over 95% packets' extra delay is less than 50us. We can see the extra packet delay by adding the buffer is minor.

V. Conclusions

Low average utilization and redundant processing capacity in today's high-end routers provide opportunities for power saving. By unbinding network interfaces and packet processing engines in line-cards, traffic is aggregated to active MBs on-demand, so we can deactivate other MBs to power off for saving power. In this paper, we first design a two-stage switch fabric to interconnect DBs and MBs in a fully sharing manner. Evaluations based on real network traffic show that GreenRouter is able to achieve considerable power saving while guaranteeing the network QoS performance. Besides, GreenRouter is easily implemented in engineering and more propitious to system on-powering maintenance.

References

[1] J. Guichard, F. L. Faucheur, and J. P. Vasseur, *Definitive MPLS network designs*. Cisco Press, 2005.

[2] K. Thompson, G. J. Miller, and R. Wilder, "Wide-area internet traffic patterns and characteristics," *IEEE Network*, vol. 11, no. 6, pp. 10–23, 1997.

[3] C. Labovitz, "What europeans do at night," http://ddos.arbornetworks.com/2009/08/what-europeans-do-at-night/.

[4] 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," *Communications Surveys & Tutorials*, vol. 13, no. 2, pp. 223–244, 2011.

[5] S. Nedevschi, L. Popa, G. Iannaccone, S. Ratnasamy, and D. Wetherall, "Reducing network energy consumption via sleeping and rate-adaptation." in *NSDI*, 2008.

[6] M. Zhang, C. Yi, B. Liu, and B. Zhang, "Greente: Power-aware traffic engineering," in *IEEE International Conference on Network Protocols (ICNP)*, 2010.

[7] H. Anand, C. Reardon, R. Subramaniyan, and A. D. George, "Ethernet adaptive link rate (alr): Analysis of a mac handshake protocol." in *LCN*, 2006.

[8] C. Gunaratne, K. Christensen, B. Nordman, and S. Suen, "Reducing the energy consumption of ethernet with adaptive link rate (alr)," *IEEE Transactions on Computers*, vol. 57, no. 4, pp. 448–461, 2008.

[9] M. Andrews, A. F. Anta, L. Zhang, and W. Zhao, "Routing for energy minimization in the speed scaling model," in *IEEE INFOCOM*, 2010.

[10] C. Gunaratne and K. Christensen, "Ethernet adaptive link rate: System design and performance evaluation," in *LCN*, 2006.

[11] M. Gupta and S. Singh, "Using low-power modes for energy conservation in ethernet lans." in *IEEE INFOCOM*, 2007.

[12] P. Reviriego, J. A. Maestro, and D. Larrabeiti, "Burst transmission for energy-efficient ethernet," *IEEE Internet Computing*, vol. 14, no. 4, pp. 50–57, 2010.

[13] S. Herbert and D. Marculescu, "Analysis of dynamic voltage/frequency scaling in chip-multiprocessors," in *ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*, 2007.

[14] M. Mandviwalla and N.-F. Tzeng, "Energy-efficient scheme for multiprocessor-based router linecards," in *International Symposium on Applications and the Internet*, 2006.

[15] A. Miyoshi, C. Lefurgy, E. Van Hensbergen, R. Rajamony, and R. Rajkumar, "Critical power slope: understanding the runtime effects of frequency scaling," in *Proceedings of the 16th international conference on Supercomputing*, 2002.

[16] L. Shang, L.-S. Peh, and N. K. Jha, "Dynamic voltage scaling with links for power optimization of interconnection networks," in *The Ninth International Symposium on High-Performance Computer Architecture*, 2003.

[17] A. Kennedy, X. Wang, Z. Liu, and B. Liu, "Low power architecture for high speed packet classification," in *Proceedings of the 4th ACM/IEEE Symposium on Architectures for Networking and Communications Systems*, 2008.

[18] A. Kennedy, X. Wang, and B. Liu, "Energy efficient packet classification hardware accelerator," in *IEEE International Symposium on Parallel and Distributed Processing(IPDPS)*, 2008.

[19] W. Meng, Y. Wang, C. Hu, K. He, J. Li, and B. Liu, "Greening the internet using multi-frequency scaling scheme," in *IEEE International Conference on Advanced Information Networking and Applications (AINA)*, 2012.

[20] M. Yamada, T. Yazaki, N. Matsuyama, and T. Hayashi, "Power efficient approach and performance control for routers," in *IEEE International Conference on Communications Workshops*, 2009.

[21] A. Bianco, F. G. Debele, and L. Giraudo, "Energy saving in distributed router architectures," in *IEEE International Conference on Communications (ICC)*, 2012.

[22] Y. Kai, Y. Wang, and B. Liu, "Greenrouter: Reducing power by innovating router," *Computer Architecture Letters*, vol. 12, no. 2, pp. 51–54, 2013.

[23] C.-S. Chang, D.-S. Lee, and Y.-S. Jou, "Load balanced birkhoff–von neumann switches, part i: one-stage buffering," *Computer Communications*, vol. 25, no. 6, pp. 611–622, 2002.

[24] W. Shi, M. H. MacGregor, and P. Gburzynski, "A scalable load balancer for forwarding internet traffic: exploiting flow-level burstiness," in *Proceedings of the 2005 ACM symposium on Architecture for networking and communications systems*, 2005.

[25] L. J. Wobker, "Power consumption in high-end routing systems," http://www.nanog.org/meetings/nanog54/presentations/Wednesday/Wobker.pdf.

[26] B. Liu, J. Lu, Y. Kai, and Y. Wang, "Enabling power-porportional routers," Technical Report, http://s-router.cs.tsinghua.edu.cn/%7elujianyuan/Enabling%20Power-proportional%20Routers.pdf.

[27] C. Walsworth, E. Aben, kc claffy, and D. Andersen, "The caida ucsd anonymized internet traces 2010," http://www.caida.org/data/passive/passive_2010_dataset.xml.

[28] ——, "The caida ucsd anonymized internet traces 2012," http://www.caida.org/data/passive/passive_2012_dataset.xml.