Power Awareness in Network Design and Routing

Joseph Chabarek*, Joel Sommers*, Paul Barford*, Cristian Estan*, David Tsiang†, Steve Wright*
*University of Wisconsin-Madison, (jpchaba,jsommers,pb,estan,swright)@cs.wisc.edu
†Cisco Systems, tsiang@cisco.com

***Abstract*—Exponential bandwidth scaling has been a fundamental driver of the growth and popularity of the Internet. However, increases in bandwidth have been accompanied by increases in power consumption, and despite sustained system design efforts to address power demand, significant technological challenges remain that threaten to slow future bandwidth growth. In this paper we describe the power and associated heat management challenges in today's routers. We advocate a broad approach to addressing this problem that includes making power-awareness a primary objective in the design and configuration of networks, and in the design and implementation of network protocols. We support our arguments by providing a case study of power demands of two standard router platforms that enables us to create a generic model for router power consumption. We apply this model in a set of target network configurations and use mixed integer optimization techniques to investigate power consumption, performance and robustness in static network design and in dynamic routing. Our results indicate the potential for significant power savings in operational networks by including power-awareness.**

I. Introduction

In the last decade, the Internet transformed from a computer network used primarily by academics into a worldwide communication medium with significant impact on the global economy. The role of the Internet will perhaps be even more important in the future, prompting Ray Ozzie, Microsoft's CTO to recently state that "We're in a new era—an era in which the Internet is at the center" [1]. The tremendous growth in the number of end users and their network connection speeds has resulted in a consistent, exponential increase in the bandwidth demand. To keep pace, Internet Service Providers have relied on similar growth in bandwidth and capacities of routers and switches. These performance improvements have been accompanied by a decrease in the cost per byte of traffic, which has further fueled the growth of the Internet by making connectivity more affordable for everyone.

Today's high performance router line cards handle most data plane traffic processing tasks with specialized ASIC hardware. Decreasing feature sizes in semiconductor technology have contributed to performance gains by allowing higher clock frequencies and design improvements such as increased parallelism. The same technology trends have also allowed for a decrease in voltage that has reduced the power per byte transmitted by half every two years as illustrated in Figure 1. However, since the rate at which line card speeds increase is greater, there has been an overall *increase* in power density. Unfortunately, the power efficiency of the underlying technology is starting to plateau, and as the power savings due to technology improvements slows down, the rate of increase for power density will accelerate. At the same time, the heat dissipation demands of routers are reaching the limits of traditional solutions based on air cooling.

This confluence of technology trends forebodes grave consequences if power consumption continues along the present trajectory. Expensive liquid cooling may soon be required for high performance routers. Increases in floor space and heat dissipation costs of new multi-chassis devices will cause increases in the cost of operating network points of presence (PoPs) and will require significant investments in new facilities. In the long term, this could lead to a slowing of the rate of decrease of the cost of carrying traffic, which may well have a chilling effect on the continued growth of the Internet.

In this paper, we examine the problem of power-awareness in wire-line networks. We argue that to combat the grim scenario described above, we must go beyond current efforts focused only on system-oriented power management. We describe two directions that we believe will lead to substantial reduction in power requirements for network devices: power-aware network design and power-aware protocol design. We describe these areas in Section III and support our arguments through a series of experiments examining the application of power-awareness in network design and routing.

We begin by measuring the power demand of two widely used routers over a range of configurations. This enables us to create a generic model for power consumption of network devices. We apply this model in a set of network topologies and associated traffic matrices to investigate power demand in network design. In optimization terms, the resulting formulation is a design problem overlaying a multicommodity network-flow problem. By formulating this problem as a mixed-integer program and solving it with the help of state-of-the-art modeling systems and optimization methods, we are able to find system configurations that minimize power consumption while preserving performance and robustness requirements. Using the lessons learned from these experiments, we assess the potential impact of power awareness in routing.

Specifically, we consider how routes might be adjusted on relatively coarse time scales such that network-wide power consumption can be minimized. The intuition is that well-known daily fluctuations in traffic load may offer opportunities to reroute traffic in order to save power. Through further application of mixed integer optimization enables us to establish a minimum power consumption profile for a set of random networks with increasing levels of connectivity. The results power savings are possible, although additional work will be required in order to realize these possibilities in operational

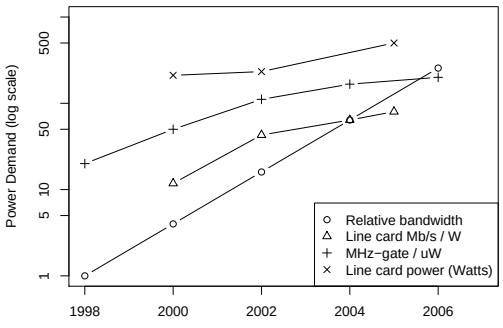

Fig. 1. Line card versus bandwidth versus power demands since 1998. NOTE that Watts and Microwatts are used in power demand to show the general trend. Data sources: IBM (MHz-gate/uW), Cisco Systems (Relative bandwidth, Line card Mb/s / Watt, Line card power)

environments.

II. Related Work

McKeown has identified increasing power density trends in routers in several talks (*e.g.,* [2]). In [3], he and his co-authors describe the use of optics in routers as a means for scaling capacity and reducing power consumption. Similarly, Minkenberg *et al.* highlight power consumption as one among a number of issues and trends in the design of packet switch devices [4], while Wassal and Hasan investigate methods for decreasing power consumption in interconnection fabrics in [5]. Our perspective is that power consumption and heat dissipation have fast become the primary issues in router system design. Furthermore, any reduction in size or computational complexity of critical components in network systems will reduce power density, and we believe all aspects of systems must be considered. One of the arguments in our paper is that resource reductions similar to [6] might be possible—especially if basic design requirements can be modified.

In [7] Gupta and Singh suggest the idea of energy conservation in Internet systems. The thesis of their paper is that components in network devices can be put to sleep (or into energy saving modes) with some changes to Internet protocols in order to save energy. Their idea of coordinated sleeping is similar to our notion of power-aware routing. They explore this idea in a wired LAN setting in [8]. We consider a more coarse-grained network design and routing approach.

There is a large literature on power-awareness in mobile ad-hoc and wireless networks. Jones *et al.* provide a useful survey of many of these techniques [9]. There are also a number of complementary studies on improving energy efficiency in operating systems for mobile devices (*e.g.,* [10]). Likewise, the computer architecture community has been concerned with energy consumption in chip design for some time and has developed a variety of methods for addressing this problem (*e.g.,* [11], [12]). Despite these advances, serious challenges remain in developing and deploying energy efficient systems and protocols in *wire-line* networks.

Optimization techniques have been applied to many different problems in communication networks. Recent work by Applegate and Cohen uses an optimization framework to investigate how changing traffic profiles can affect network utilization [13]. Our approach to investigating power-awareness is similar in that we consider both network topologies and traffic matrices. Zhang *et al.* use an optimization framework to consider the problem of how to generate a set of routes that will provide acceptable performance over a range of expected traffic matrices [14]. Aspects of our general approach and analysis methodology are similar to that work. Finally, optimization methods have also been applied to network design and provisioning problems. A canonical reference is [15]. To the best of our knowledge, there is no prior work that considers power-awareness in wire-line network design.

III. Power Aware Design Space

We argue that as we near the basic limits of today's semiconductor technology, a much more comprehensive treatment of the power density problem in routers is required. Three areas that we believe hold the most promise include power-aware system design, power-aware network design and power-aware protocols. While power-aware system design is not the focus of our work, understanding the methods typically used by router manufacturers to reduce power consumption and mitigate heat informs our approach to power-aware network and protocol design. In this section, we provide an overview of each of these three areas.

A. Power-aware System Design

Bandwidth growth has been based to a large extent on new developments in CMOS technology. Standard techniques for power efficient design in router ASICs include clock gating, process-specific supply voltages, and reducing supply voltages. But, as Figure 1 shows, the rate of CMOS power efficiency improvement is slowing, which has led to innovations in router system design to address the associated heat dissipation requirements. Two additional methods for keeping up with demand for exponential bandwidth growth within the constraints of traditional air cooling include:

1) Multi-Chassis Systems: Multi-chassis routers allow separate physical components to be clustered together to form a single logical router. A common architecture for a multi-chassis router consists of several line card chassis connected to a non-blocking scalable switch fabric chassis. Multiple chassis solve the bandwidth scaling problem by providing a growth path that does not rely on increasing the bandwidth density and power density. Although the aggregate power consumption increases, the heat load is spread over a large physical area which allows existing air-cooling techniques to be used, at the cost of requiring additional physical space in a PoP.

2) Alternative Systems: The optical switch has long been considered the primary candidate for replacing the electronic router. Pure optical switches have the allure of being able to provide terabits of bandwidth at much lower power dissipation than electronic switches. Furthermore, they can be almost entirely bit-rate independent because of their ability to switch a broad spectrum of light (hence many wavelengths may be

switched simultaneously). The practical use of optical switches has been hampered over the years by several problems. First, the number of ports has been limited by technology to less than 100. This limitation makes them suitable only for the core of the network. However, use in the core of the network requires long-haul capability which in turn requires complex optical network engineering to achieve long transmission distances. A second problem is that significant optical buffering is not currently feasible. Nevertheless, optical technology continues to evolve and may have an important impact on reducing power consumption in the future.

B. Power-aware Network Design

Power-awareness in network design offers the opportunity to deploy routers over a set of PoPs such that the aggregate power demand is minimized while requirements for robustness and performance are satisfied. We envision two approaches towards this goal. First, there are likely to be multiple router-level network topologies that can satisfy a given set of capacity, robustness and power consumption design objectives. Our quantitative evaluation from Section VI demonstrates that being aware of power consumption when designing network topologies can result in significant power reductions. Second, the network can be designed such that power-hungry packet processing operations are limited to a subset of the routers.

Current network design, configuration and management practices are based on deploying and maintaining infrastructures that are extremely reliable, provide performance that enables competitive service level agreements and offer a set of features and services that are attractive to a broad range of customers. To accomplish these goals, network architects typically build infrastructures that are densely interconnected with many redundant paths using state-of-the-art high bandwidth routers in the core, lower bandwidth but high connection density distribution routers around the core and even lower bandwidth access routers and switches at the periphery. It is important to note that equipment manufacturers have traditionally provided the most functionality in core routers, which is likely to have been a consequence of the large processing capacity of these systems as well as competitive and economic factors. As a result, network design, configuration and management practices have been organized around taking advantage of these capabilities in the core and to a lesser extent on the edges.

With respect to power-aware network design, the long term objective is to replace power-hungry systems in the core with lower power systems that still provide required reliability and performance. By understanding the power demands of today's routers and switches under different configurations and traffic loads, ISPs have an opportunity to develop power and heat budgets for their networks that can save energy costs and potentially reduce equipment footprints in PoPs.

C. Power-aware Protocols

The final dimension of power-awareness that we advocate is in the design and implementation of network protocols. This notion fits quite well with the traditional end-to-end arguments [16], and perhaps provides additional perspective beyond performance considerations. While power-aware protocols have been investigated for some time in the wireless context, we believe there are many opportunities for valuable developments in wire-line networks.

As described in Section II, perhaps the most basic notion of power-aware protocols include mechanisms for putting components to sleep. Development of new data link and routing protocols could, *(i)* make traffic profiles more efficient (*e.g.,* auto-negotiate PPS rate or minimum packet size), *(ii)* enable portions of a line card to be turned off (if certain features or ports are not in use), or *(iii)* enable entire line cards to enter a hibernation state (which could be an objective of a power-aware routing protocol). We consider the potential impact of power-aware routing protocols in Section VI, but leave development of such a protocol for further work.

IV. Benchmarking Router Power Consumption

We begin our investigation of power-awareness in network design and routing by conducting an empirical study of power consumption in two widely used routers. We begin by measuring the gross characteristics of power consumption through experiments with different combinations of line cards in each chassis. We then measure the more subtle aspects of power consumption on a single device through experiments over a range of (soft) configurations and operating conditions. We use the combined set of measurements in formulating a general model for router power consumption.

A. Idle Chassis/Line card Combinations

In our first set of experiments, we measured the power consumption of two *idle* router chassis with different combinations of line cards installed. Our target platforms were a Cisco GSR 12008 and a Cisco 7507. The 7507 is a seven slot device that can accommodate up to 1 Gb/s per slot and is designed for operation at a network edge. The 12008 is a ten slot device (two are dedicated to the switching fabric) that can accommodate up to 4 Gb/s per slot and is designed for operation in a network core. The line card configurations we used for these two routers is shown in Table I. While these configurations were chosen to be representative of a range of common networking technologies, the hardware is produced by one vendor so the configurations cannot be a guarantee of the general applicability of the benchmarks.

Our power measurement device was a Fluke 189 digital multimeter equipped with an i200s AC current clamp [17]. During our experiments this measurement clamp was attached to the power cable of the router. This set up enabled us to measure system-wide power consumption in our experiments.

Experiments began by setting up a specific router/line-card configuration. Cables were removed from all interface ports on installed line cards, and line cards that were not used in a given test configuration were removed from the chassis. Each test began by turning on the router and then waiting for sufficient time for it to initialize. We then measured power consumption

TABLE I
ROUTER/LINE-CARD CONFIGURATIONS USED IN POWER CONSUMPTION BENCHMARKING EXPERIMENTS.

Chassis Slot	Line Card	Abbreviation
0	Empty	None
1	4 port GE line card	4GE
2	4 port OC-3/POS line card	OC-3
3	1 port OC-48/POS line card	OC-48
4	10 Gb/s Switching fabric	CSC
5	10 Gb/s Switching fabric	CSC
6	4 port OC-12/POS line card	OC-12
7	4 port GE line card	4GE
8	Route Processor	RP
9	Empty	None

(a) Cisco GSR 12008 configuration.

Chassis Slot	Line Card	Abbreviation
0	1 port GE line card	GE
1	1 port FE line card	FE
2	Route Processor	RP
3	Empty	None
4	1 port FE line card	FE
5	1 port DS1 line card	DS1
6	1 port FE line card	FE

(b) Cisco 7507 configuration.

once per second over 200 seconds for the GSR and 400 seconds for the 7507. The longer measurement period for the 7507 was due to a higher degree of variability during tests. The power consumption measurements were then averaged over the test period.

The measurement results for the GSR are shown in Figure 2 and the results for the 7507 are shown in Figure 3. For each configuration we observe that the base system (*i.e.*, chassis plus a router processor for the 7507, chassis plus a router processor plus switching fabric for the GSR) consumes more than half the maximum observed power consumption for any configuration. For the GSR, the base system consumes approximately 430 Watts, and the 7507 base system consumes approximately 210 Watts.

As line cards beyond those required for the base system are installed, the level of power consumption increases in discrete steps depending on the line card type. For example, the one-port OC-48 card in the GSR consumes an additional 70 Watts, while the one-port FE card in the 7507 consumes an additional 25 Watts. Table II shows average power consumption for several other line cards. For the GSR, we note that the base configuration plus the one-port OC-48 line card consumes about the same power as the base configuration with a four-port OC-12 line card. These cards have the same internal bandwidth (2.5 Gb/s) and have similar feature capabilities (each is designated "engine 2"). Overall, the base system is the largest consumer of power on either system, even in the most dense configurations we examined.

These results lead directly to the conclusion that from a power-aware perspective, it is best to minimize the number of chassis that are powered at a given PoP, and to maximize the number of line cards per chassis.

B. Effects of Configuration and Operating Conditions

To understand how configuration and operating conditions affect power consumption, we set up a testbed consisting of

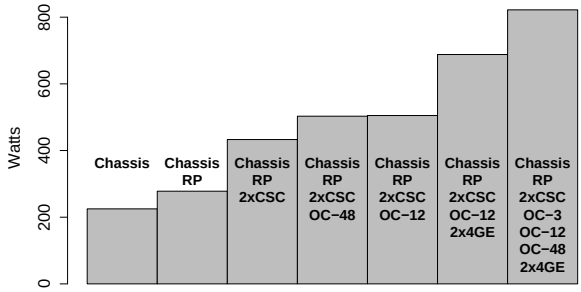

Fig. 2. Power consumption for different configurations of the GSR. Each configuration is labeled according to the installed line cards per the abbreviations in Table IIa.

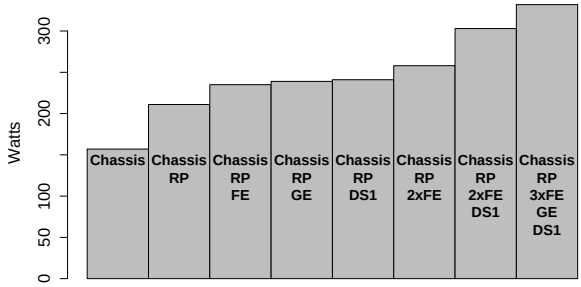

Fig. 3. Power consumption for different configurations of the 7507. Each configuration is labeled according to the installed line cards per the abbreviations in Table IIb.

commodity workstation end hosts and commercial IP routing systems configured in a dumbbell-like topology as depicted in Figure 4. We used 20 workstations for traffic generation, each of which had a Pentium 4 processor running at 2 GHz or better, at least 1 GB RAM, and an Intel Pro/1000 network interface card. Each host was configured to run either FreeBSD 5.4 or Linux 2.6. End host's packet traffic was aggregated using Cisco 6500 routers and flowed through two Cisco GSR 12008s. Three parallel Gigabit Ethernet links connected the 6500s to the GSRs, and the GSRs were connected via a OC-48 link.

The focus of our evaluation was a 4-port Gigabit Ethernet engine 3 line card in the GSR chassis, *i.e.*, the ingress line card for the device under test (DUT) in Figure 4. This line card has an internal capacity of approximately 2.5 Gb/s at 4

TABLE II
ROUTER LINE CARD POWER CONSUMPTION (NO TRAFFIC).

Line card type	Power (watts)
4 port GE	92
4 port OC-12/POS	72
1 port OC-48/POS	70

(a) GSR line card power consumption.

Line card type	Power (watts)
1 port Fast Ethernet	26
1 port Gigabit Ethernet	30
1 port 1.544 Mb/s DS1	49

(b) 7507 line card power consumption.

million packets per second (Mp/s).

The Harpoon traffic generator was used for creating constant bit rate UDP traffic, as well as self-similar TCP traffic [18]. We used a class B network (2^{16} addresses) for source addresses, and another class B network for destination addresses. Each experiment was run for 10 minutes. We configured the multimeter to store its average value every second, thus providing 600 data points per experiment (the first and last 30 seconds are omitted). We report first order statistics on power consumed using the digital multimeter described above for the following configurations:

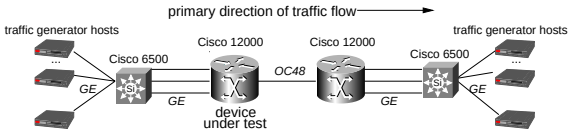

Fig. 4. Laboratory testbed.

1) Baseline/Idle. This experiment establishes a power use baseline against which other experiments will be compared. No traffic flows through the DUT.
2) Data plane/Small packets. This experiment is designed to measure power use when the data plane switches small packets. The forwarding table at the DUT consists of about 1000 entries. Traffic consists of constant bit rate, 100 byte packets at just under the 2.5Gb/s capacity of the 4-port GE line card. This setup results in approximately 2 Mp/s of traffic through the line card.
3) Data plane/Medium packets. Same as #2, but with 576 byte packets. This setup results in approximately 540 Kp/s through the GE line card.
4) Data plane/Large packets. Same as #2, but with 1500 byte packets. This setup results in approximately 200 Kp/s through the GE line card.
5) Data plane/Forwarding table. This experiment is designed to measure power use when a large forwarding table (about 32,000 entries) is configured. Traffic is the same as #3.
6) Data plane/Bursty traffic. This experiment is designed to establish baseline power use for self-similar TCP traffic. Harpoon is configured to produce, on average, 75% of the 2.5Gb/s capacity of the 4-port GE line card.
7) Features/ACLs. This experiment is designed to measure power use when a 1000-entry ACL is configured. Configuration is otherwise the same as #3.
8) Features/uRPF. This experiment is designed to measure power use when unicast reverse path forwarding is used. Configuration is otherwise the same as #3.
9) Features/AQM: This experiment is designed to measure power use when RED (default configuration) is turned on. Traffic is the same as #6.
10) Measurement/NetFlow. This experiment is designed to measure power use when NetFlow (default configuration) is enabled. Traffic is the same as #6.
11) Control plane/OSPF updates. This experiment is designed to measure power use during OSPF routing updates. The configuration is the same as #3, plus OSPF link state advertisement traffic that arrives every 15 seconds resulting in forwarding table updates.

Figure 5 shows box-and-whiskers plots for the baseline/idle scenario, and for three packet sizes for constant-bit rate scenarios. The interquartile range with median is shown, along with the range of values (including any outliers). The important features of this plot are, *(i)* the difference between the idle scenario and the scenarios where packet traffic is present; *(ii)* the increasing trend in power consumption as packets get smaller (or, alternatively, as packet rate increases); *(iii)* the absolute difference between the idle state and the scenarios with packet traffic is about 20 watts.

Figure 6 shows box-and-whiskers plots for the baseline/idle scenario along with various features enabled for constant bit rate traffic with medium-sized packets. While power consumed is similar in each of the non-idle cases, the highest consumer is the uRPF scenario. Interestingly, the power consumed in the case with a large forwarding table versus the baseline 1,000-entry table is actually less. Also, while the median power consumed in the ACL scenario is about the same as the baseline medium-sized packet scenario, the variability is somewhat higher.

Finally, Figure 7 shows box-and-whiskers plots for the baseline/idle scenario along with the baseline self-similar traffic and self-similar traffic with NetFlow and RED scenarios. In each of the self-similar traffic scenarios, the power consumed is about the same as for the large-sized packet constant bit rate scenario. For each of the self-similar scenarios, the variability is higher than for other scenarios.

The maximum variation in power use in our experiments, about 20 watts, is for a two line card configuration on the GSR (the ingress GE card, and the egress OC-48 card). With a fully loaded 12008 chassis, we extrapolate this difference to be between 150 and 200 watts or about 10% of the rated maximum. Clearly, these effects are less significant than those related to chassis/line card configurations, but cannot be discounted. Furthermore, the relatively high baseline of about 755 watts indicates that there may be opportunities for low power/hibernation modes for different components on this system or for designing their power consumption to be tied more directly to load.

C. A General Model for Router Power Consumption

The benchmarking results reported above lead to the following generalized model for router power consumption:

$$PC(X) = CC(x_0) + \sum_{i=0}^{N} (TP(x_{i0}, x_{i1}) + LCC(x_{i1})) \quad (1)$$

The power consumption PC of a router is determined by its configuration and current use. The vector X defines the chassis type of the device, the installed line cards and the configuration and traffic profile of the device. The function $CC(x_0)$ returns the power consumption of a particular chassis

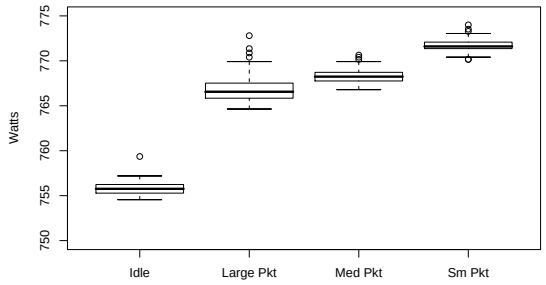

Fig. 5. Results for constant bit rate UDP traffic with different packet sizes at about 2.5 Gb/s, along with idle baseline.

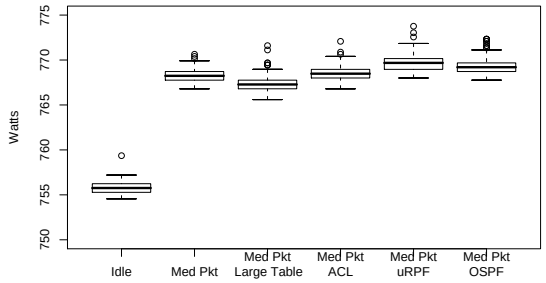

Fig. 6. Results for constant bit rate UDP traffic with medium packet size with different features enabled, along with idle baseline.

type, N is the number of line cards that are active, $TP(x_{i0})$ is a scaling factor corresponding to the traffic utilization on the router, and $LCC(x_{i1})$ gives the cost of the line card in a base configuration. The cost of traffic can be any function and is dependent on the configuration of the router and the amount of traffic. We use this model in the following section to formulate our optimization problem for power-aware network design.

V. Optimizing Power Consumption

We use a flexible optimization framework to allocate resources in target networks in a power-aware fashion. To promote power-awareness, we focus on allocation of line cards and chassis over target networks. While traffic was shown in Section IV to have some impact on the power consumption of a line card, the costs of powering the chassis and the line cards themselves dominate the overall power profile of a network, and we ignore the TP term in Equation 1. In addition to standard network hardware functionality, for coarse-grained routing analysis, we assume that service providers can dynamically power on/off line cards and chassis. In this section, we describe how the network design problem can be formulated as a multicommodity flow problem with design variables that indicate the configuration (line cards and chassis) at each node. The number of integer and binary variables in the formulation is not exceptionally large, but the very large number of commodities (each commodity representing traffic between an origin-destination pair of nodes), together with the associated flow-balance constraints, makes the full problem quite large. In general, mixed-integer programs are known to be NP-hard,

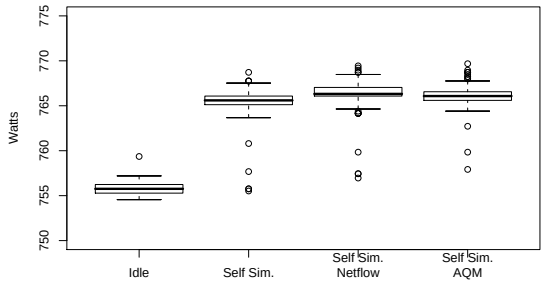

Fig. 7. Results for self-similar traffic at an average offered load of 75% of the GE line card capacity of 2.5 Gb/s with NetFlow and RED enabled, along with idle baseline.

but sophisticated heuristics (in particular, branch-and-bound techniques and cut generation techniques) can be used to solve many instances of these problems in a reasonable amount of time, especially when guided by a user with problem-specific knowledge.

A. GAMS Based Optimization

Our model takes as input a network annotated with OSPF link weights, a traffic matrix for this network, and line card and chassis options for provisioning each node of the network. As output, the optimization process determines how each node should be provisioned in order to minimize network-wide power consumption. It also discovers the multipath routing to be used for traffic from each origin-destination (O-D) pair in the traffic matrix. In optimization terms, our model is a mixed-integer resource allocation problem with multicommodity flow constraints.

We introduce some notation to describe the parameters and variables in the model and the relationships between them.

1) Network topology: a set of nodes $1, 2, \ldots, N$ (for which we typically use the generic indices i and j) and an arc set *arc*, in which a directed arc from node i to node j is denoted by $i \rightarrow j$. Weights on the arcs correspond to OSPF link weights.
2) Parameters that represent possible hardware configurations. These include the available chassis $c = 1, 2, \ldots, C$ and line card models $l = 1, 2, \ldots, L$. For each chassis, $cost_C(c)$ is the operating cost in Watts. We denote the number of line cards of type l that can be accommodated in chassis c as $q_{l,c}$. For each line card, $cost_L(c)$ denotes the cost in Watts for operating the line card. In addition, for each line card l an effective maximum card throughput T_l in bits per second is assigned, along with the number of ports available P_l for a line card.
3) Traffic. We use the standard multicommodity flow problem formulation where for each O-D pair a commodity is allocated. Given a specified flow $d_{O,D}$ from node O to node D, we set the supply of commodity (O, D) at node O to be $d_{O,D}$, and the supply at node D to be $-d_{O,D}$, with a supply of zero at all other nodes. The amount of flow from O to D that is routed along the arc $i \rightarrow j$ is denoted by $f_{i \rightarrow j}(O, D)$.

Our model determines the flow routing variables $f_{i\to j}(O,D)$ along with the number $m_{l,c}(i)$ of line cards of type l that are allocated to chassis c at node i, and the number $n_c(i)$ of chassis of type c that have been allocated at node i. (Clearly, $m_{l,c}(i)$ and $n_c(i)$ are integer variables.) Our formulation leverages standard multicommodity flow constraints, plus a number of constraints that are specific to allocating devices. These constraints include:

1) There must be a sufficient number of chassis at each node to accommodate the line cards selected by the model:

$$\sum_{l:q_{l,c}>0} \frac{m_{l,c}(i)}{q_{l,c}} \le n_c(i), \;\; i=1,2,\ldots,N, \; c=1,2,\ldots,C. \tag{2}$$

2) The total amount of traffic being routed through this node can be accommodated by the selected line cards:

$$\begin{aligned}\textstyle\sum_{i\to j\in arcs} \sum_{(O,D)} f_{i\to j}(O,D) + \sum_O d_{O,i} \\ \le \textstyle\sum_{c=1}^{C} \sum_{l=1}^{L} m_{l,c}(i) * T_l,\end{aligned} \tag{3}$$

for $i=1,2,\ldots,N$. Note that we define the amount of traffic to be handled by the line cards at node i to be the amount of traffic flowing out of the node, plus the amount of traffic for which node i is the destination. The later term is obtained by summing the quantities $d_{O,i}$ over all possible origin nodes O. By avoiding any double counting of the traffic through the node we are modeling the full duplex properties of our line cards.

3) The number of arcs either entering or leaving node i that have nonzero flow must not exceed the total number of ports available at the node:

$$\begin{aligned}count_{j\neq i}\Big(\sum_{(O,D)} f_{i\to j}(O,D) \neq 0 \vee \\ \sum_{(O,D)} f_{j\to i}(O,D) \neq 0\Big) \\ \le \sum_{c=1}^{C}\sum_{l=1}^{L} m_{l,c}(i) P_l,\end{aligned} \tag{4}$$

for $i=1,2,\ldots,N$. This constraint is implemented by introducing a binary variable $b_{i\to j}$ for each arc $i\to j$, which takes the value zero when no flow is routed along the arc $i\to j$, and 1 otherwise. These settings are enforced by applying the constraints

$$\sum_{(O,D)} f_{i\to j}(O,D) \le b_{i\to j} C_{i\to j},$$

where $C_{i\to j}$ is the capacity of the link $i\to j$. A second set of binary variables $\bar{b}_{i,j}$ can be defined along with appropriate constraints, to ensure that $\bar{b}_{i,j}=1$ if either $b_{i\to j}=1$ or $b_{j\to i}=1$, and $\bar{b}_{i,j}=0$ otherwise. Constraint (5) is then formulated as follows:

$$\sum_{j\neq i} \bar{b}_{i,j} \le \sum_{c=1}^{C}\sum_{l=1}^{L} m_{l,c}(i) P_l, \;\; i=1,2,\ldots,N.$$

We considered a number of power-aware objective functions. A function based solely on minimizing the cost of provisioning the network is as follows:

$$\sum_{i=1}^{N}\sum_{c=1}^{C}\left[\left(\sum_{l=1}^{L} cost_L(l) m_{l,c}(i)\right) + cost_C(c) n_c(i)\right]. \tag{5}$$

We implemented this model using the General Algebraic Modeling System (GAMS) [19], which is a high-level language for optimization model development. GAMS translates such models into a form that can be recognized by codes for solving optimization problems, such as the CPLEX [20]. Once the optimization code solves the problem, GAMS interprets the solution in terms of the user-specified model.

CPLEX implements a state-of-the-art branch-and-cut solver for mixed-integer programming, which combines a branch-and-bound strategy for the integer variables with the generation of additional constraints ("cuts") that exclude regions of the feasible space that are determined not to contain the solution. However, our problems are large and complex even by the standards of the best optimization software available today, and default settings for GAMS and CPLEX are in some instances inadequate. We need to set various options (such as instructing the solver to search for the network structure, to generate cuts of various types) and define branch ordering for the integer variables in order to obtain solutions in a reasonable amount of time.

B. Strawman Configurations

We do not have ground truth power consumption measurements for any live network, or the number and type of hardware devices deployed in live networks (this information is highly proprietary). We therefore endeavor to create realistic test network configurations that can be used to assess the impact of power-awareness in network design and routing. To do this we choose the chassis and line card with the smallest power usage per bit when a chassis N is filled to capacity with line card of type M which we proceed to allocate uniformly across the network. This allocation is done in a bin packing fashion, where the traffic assigned to each node is computed ahead of time with a shortest multi-path objective function given OSPF weights.

C. Test Networks and Traffic Matrices

Our power consumption analysis uses annotated networks with inferred weights and link latencies provided by the Rocketfuel project [21]. The two largest annotated networks in this database are beyond the computational capabilities of our mixed-integer optimization formulation. To generate a synthetic traffic matrix from the graphs provided by Rocketfuel, we use a gravity model [22] in a manner similar to Applegate and Cohen [13] where the inferred link weights are used to calculate approximate bandwidths of each link at a node. These bandwidths are used in the gravity model to derive the proportion of traffic for which each node (PoP) is responsible and therefore how much traffic it sends to all the other nodes. In addition, we consider three simple graphs with a small,

constant number of nodes and varying numbers of directed edges. The random graphs were created with the Brite network generator [23] utilizing the Waxman method.

TABLE III
TEST NETWORKS USED IN OUR POWER CONSUMPTION STUDY. THE DESCRIPTION FOR THE ROCKETFUEL NETWORKS INCLUDE THE ISP NAME AND AS NUMBER, WHILE THE RANDOM GRAPHS GENERATED BY BRITE HAVE LABELS BEGINNING WITH RAND.

Description	Nodes	Directed Edges
Telstra (AS) 1221	7	18
Ebone (AS) 1755	18	66
Exodus (AS) 3967	21	72
Abovenet (AS) 6461	17	74
Rand24	12	24
Rand48	12	48
Rand134	12	134

VI. EXPERIMENTAL RESULTS

Using the GAMS framework and the test networks listed above, in this section we investigate power consumption in network design and routing. The network design problem considers how different chassis/line card configurations might be deployed in a network such that provisioning requirements are satisfied while power consumption is minimized. In our experiments, we consider a range of possible chassis/line card configurations based on our benchmarks presented in Section IV, we vary provisioning requirements by scaling the traffic loads in the test networks. The routing problem considers how traffic flows might be altered in order to put line cards and/or chassis to sleep during periods of low utilization. In this case the chassis/line card deployment is fixed, and traffic load is varied. We stress that the objective of our study is to expose the relationship between power consumption, network configuration and provisioning and that the specific values for power consumption presented below are meaningful only in a relative sense.

Traffic is scaled for each origin-destination pair in our traffic matrix using a simple linear scaling factor. This scaling was necessary because the compacted graphs removed redundant links between and within nodes, and we want to analyze networks over a range of demand where demand can also relate to provisioning requirements. As the scaling factor increases, the solvers use the traffic matrix and basic topology to allocate an appropriate number of chassis and line cards at each node. Multiple line cards are allocated for one link when necessary for the capacity constraint. Note that a link may not be specified between two nodes if there is no corresponding connection in the basic topology.

Our first network design analysis uses a model that only includes one instance of a chassis (the GSR) and one instance of a line card (the OC-48) and we allow 10 line cards per chassis. Limiting the design space enables the model to converge relatively quickly (two hour running time on a high power workstation in the worst case). The relatively high capacity of the line card meant that scaling factors had to be varied over a wide range in order to observe changes in power consumption. Figure 8 shows the power consumption

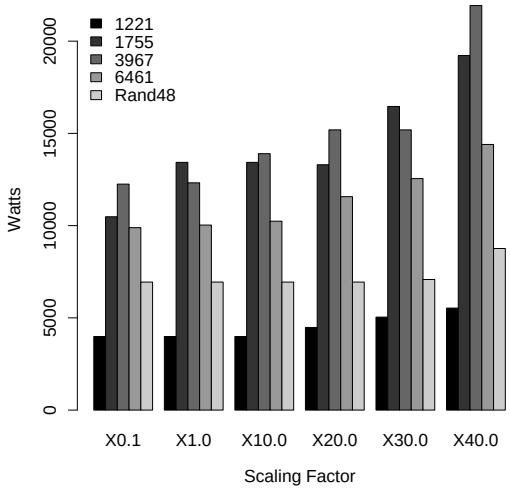

Fig. 8. Power consumption for the test networks described in Table III. The linear scaling factor is based on initial traffic loads derived using the gravity model as described in [13].

for five of the test networks. All optimization results are within 10% of optimal unless otherwise stated. The graph shows a relatively wide range of power consumption based on provisioning requirements. In most cases, as the scaling factor increases, power consumption increases in a step function-like manner related directly to additional line cards and chassis. The reason that power consumption for a scaling factor of 0.1 is the same as for 1.0 is due to our choice of line card. Had we used a lower capacity line card, lower power consumption for fractional scaling factors would have been evident.

We conducted many other experiments that relaxed some of the constraints with respect to line cards per chassis, chassis types and line card types. Space limitations prevent a detailed description. However, the important observation is that minimum power consumption coincides with chassis' that can accommodate large numbers of line cards and line card capacities that closely match demand.

In order to assess the potential impact of power awareness in the routing context, we focused our attention on the random networks which have a range of connectivity. The extra links provide redundant paths between nodes, thereby increasing robustness. They also provide an opportunity for power savings when demand on the links falls below a minimum threshold. In these instances, our solver will, in effect, shunt traffic to an alternate path in order to minimize the number of line cards and chassis in use. We first find the power use in each graph using a specific chassis/line card combination and using simple shortest path routing and no power-awareness. This strawman serves as the baseline for comparison for analyses that include power-awareness. Power savings based on the use of two different line cards are shown in Table IV. In all cases the solutions are within 11% of optimal.

The drastic improvement as the number of links increases is due in large part to the fact that the OC-48 card contains only

TABLE IV
POWER SAVINGS VERSUS SHORTEST PATH/NON-POWER-AWARE ROUTING IN THE RANDOM NETWORKS FOR TWO DIFFERENT CHASSIS/LINE CARD CONFIGURATIONS.

Network	Savings(%)
Rand24	2
Rand48	19
Rand134	65

(a) Random graph provisioning using the GSR/OC-48.

Network	Savings(%)
Rand24	2
Rand48	2
Rand134	11

(b) Random graph provisioning using the GSR/OC-12.

one ingress/egress port. Lesser effects were observed with the 4-port OC-12 line card. With a higher number of ports the cost for additional connectivity is zero as long as the number of ports needed does not require additional line cards to be allocated.

We believe that these results highlight the potential for power-aware routing protocols. For example, the heuristic method for generating routing tables for multiple traffic matrices proposed by Zhang *et al.* in [14] could be augmented with power-awareness. In this case, routes would be calculated subject to power consumption constraints. The likely outcome would be that some paths would probably not be shortest, but the resulting power savings could be substantial.

VII. CONCLUSION

Power demands in next generation networking equipment present a fundamental challenge to continued bandwidth scaling in the Internet. Relying solely on system design techniques to limit the power consumption of high speed equipment will likely not be enough to avoid expensive heat dissipation solutions such as liquid cooling. To address this problem, we advocate power-awareness in the design, configuration and management of networks, and in the design and implementation of protocols used in wire-line networks. Some of these approaches may result in direct power savings in the short term, while others can have long-term indirect effects by changing the functional requirements for routers. In this paper we present an investigation of the potential savings achievable through power-aware network design and routing. We conducted a measurement study of the power consumption of various configurations of widely used core and edge routers. We use these results to create a general model for router power consumption. Using this model along with mixed integer optimization techniques, we explore the potential impact of power-awareness in a set of example networks. Our results indicate that power consumption can vary by as much as an order of magnitude indicating that there may be substantial opportunities for reducing power consumption in the short term. In the future, we plan to investigate practical power-aware traffic engineering and network design methods, and to investigate modifications in network protocols that will reduce power demand.

ACKNOWLEDGMENT

The authors would like to thank Cisco Systems for their support of this work. This work was also supported in part by NSF grants CNS-0347252, CNS-0646256, CNS-0627102. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Cisco or the NSF.

REFERENCES

[1] R. Waters, "Microsoft's Ozzie Declares End to PC Era," Financial Times, July 2006.
[2] N. McKeown, "Scaling Routers Using Optics," http://yuba.stanford.edu/~nickm/talks, October 2003.
[3] I. Keslassy, S. Chuang, K. Yu, D. Miller, M. Horowitz, O. Solgaard, and N. McKeown, "Scaling internet routers using optics," in *Proceedings of ACM SIGCOMM '03*, Karlsruhe, Germany, August 2003.
[4] C. Minkenberg, R. Luijten, F. Abel, W. Denzel, and M. Gusat, "Current Issues in Packet Switch Design," in *Proceedings of ACM/USENIX HotNets '02*, Princeton, NJ, October 2002.
[5] A. Wassal and M. Hasan, "Low-power System-level Design of VLSI Packet Switching Fabrics," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 20, no. 6, June 2001.
[6] G. Appenzeller, I. Keslassy, and N. McKeown, "Sizing router buffers," in *Proceedings of ACM SIGCOMM '04*, Portland, OR, August 2004.
[7] M. Gupta and S. Singh, "Greening of the Internet," in *Proceedings of ACM SIGCOMM '03*, Karlsruhe, Germany, August 2003.
[8] ——, "Energy conservation with low power modes in Ethernet LAN environments," in *Proceedings of IEEE INFOCOM (minisymposium)*, Anchorage, Alaska, May 2007.
[9] C. Jones, M. Sivalingam, P. Agrawal, and J. Chen, "A Survey of Energy Efficient Network Protocols for Wireless Networks ," *Wireless Networks*, vol. 7, no. 4, pp. 343–358, July 2001.
[10] A. Vahdat, A. Lebeck, and C. Schlatter-Ellis, "Every Joule is Precious: the Case for Revisiting Operating System Design for Energy Efficiency," in *Proceedings of the 9th ACM SIGOPS European Workshop*, Kolding, Denmark, June 2000.
[11] V. Raghunathan, M. Srivastava, and R. Gupta, "A Survey of Techniques for Energy Efficient On-Chip Communication," in *Proceedings of Design Automation Conference '03*, Anaheim, CA, June 2003.
[12] T. Pering, T. Burd, and R. Bordersen, "The Simulation and Evaluation of Dynamic Voltage Scaling Algorithms," in *Proceedings of the International Symposium on Low Power Electronics and Design*, Monterey, CA, August 1998.
[13] D. Applegate and E. Cohen, "Making Intra-Domain Routing Robust to Changing and Uncertain Traffic Demands: Understanding Fundamental Tradeoffs," in *Proceedings of ACM SIGCOMM '03*, Karlsruhe, Germany, August 2003.
[14] C. Zhang, J. Kurose, D. Towsley, Z. Ge, and Y. Liu, "Optimal Routing with Multiple Traffic Matricies Tradeoff Between Average Case and Worst Case Performance," in *Proceedings of IEEE ICNP '05*, Boston,MA, November 2005.
[15] M. Pioro and D. Medhi, *Routing, Flow and Capacity Design in Communication and Computer Networks*. San Francisco, CA: Morgan Kaufmann, 2004.
[16] J. Saltzer, D. Reed, and D. Clark, "End-to-end Arguments in System Design," *ACM Transactions on Computer Systems*, vol. 2, no. 4, November 1984.
[17] "Fluke electronics," http://www.fluke.com/, 2007.
[18] J. Sommers and P. Barford, "Self-configuring network traffic generation," in *Proceedings of ACM SIGCOMM Internet Measurement Conference '04*, 2004.
[19] "The General Algebraic Modeling System (GAMS)," http://www.gams.com/, 2007.
[20] "ILOG CPLEX: High-performance software for mathematical programming and optimization," http://www.ilog.com/products/cplex/, 2007.
[21] "Rocketfuel: An isp topology mapping engine," http://www.cs.washington.edu/research/networking/rocketfuel/, 2007.
[22] M. Roughan, A. Greenberg, C. Kalmanek, M. Rumsewicz, J. Yates, and Y. Zhang, "Experience in Measuring Backbone Traffic Variability: Models, Metrics, Measurements and Meaning," in *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurement*, Marseille, France, 2002.
[23] "The Brite Network Topology Generator," http://www.cs.bu.edu/brite/, 2001.