An Intermittent Energy Internet Architecture

Barath Raghavan
ICSI

David Irwin
UMass

Jeannie Albrecht*
Williams College

Justin Ma
UC Berkeley

Adam Streed
UC San Diego

ABSTRACT

We examine how to re-design the Internet for an energy-constrained future powered by diffuse, intermittent, and expensive power sources. We consider the types of constraints this might place upon the Internet architecture and the manner in which important network components can function in this new environment. We then attempt to chart a path forward for future research.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design

General Terms

Design, Economics, Management

Keywords

Internet Architecture, Energy

1. INTRODUCTION

If we had a tenth the energy to run the Internet, and if that energy were ten times as volatile, how might we rethink the Internet's design? Recent studies indicate that energy costs will rise substantially over the coming decades, due primarily to the increasing difficulty of locating and extracting fossil fuels [12, 27]. Rising energy costs, increasing demands [6], ecological limits [20], and economic consequences [10] will force a fundamental change in the world's energy infrastructure and will likely include significant deployment of intermittent renewable energy sources [13]—sources that are likely to deliver less energy less reliably than we have come to expect—all during a period of economic turmoil.

In response to this future [26], we suggest the study of a simple question: how might we redesign the Internet's architecture to use intermittent energy sources? Although the networking literature is replete with studies on scaling up the Internet and the Internet architecture, little attention has been paid to re-scaling it to meet new constraints. We believe that energy uncertainty warrants a new perspective in networking research and a focus on Internet design for these possible conditions.

In this paper, we assume a) that today's energy infrastructure will transition to using primarily intermittent and diffuse sources of low-carbon energy; further, since we expect that energy supply will be significantly less than its potential demand, we assume b) that the electric grid will be unable to hide the volatility of these energy sources from points of consumption. These assumptions are debatable; it is possible that in some regions sources of energy will be predominantly nuclear or hydroelectric, and as a result more stable and plentiful. Similarly, it is possible that in some regions smart grid deployment proceeds rapidly and is able to manage a diverse energy portfolio in a manner that ensures always-on electricity. We are not claiming that our assumptions will be true, only that we believe that as part of a long-term research agenda it is important to consider the implications if they were true. The risk of ignoring these constraints is large, whereas exploring them may be fruitful even if the energy picture has a more cornucopian outcome.

Researchers have studied the issue of reducing the power consumption of networks in order to reduce cost, but there has been little work on coping with fluctuations and shortages. Although a complete redesign of the Internet is beyond the scope of this paper, our modest goal is to consider how the above constraints will affect many of the Internet's essential components and then to chart a course forward. Thus we begin by enumerating (Section 2) constraints a future Internet might face. We then use a case study (Section 3) to consider the Internet's components including routing, transport, addressing, clouds, link/physical technology, and applications; for each we consider how these constraints affect their design. Finally we discuss next steps (Section 4) and related work.

2. DESIGN CONSTRAINTS

Since systems research is an exercise in producing the best possible design given a set of design constraints, in this section we examine the constraints under which we envision a possible future Internet operating. While we believe that we will face at least some of these constraints in the future, our list is *not* a prediction but rather the setting for a design exercise. First, we briefly consider the premises that guide our thinking.

We envision an Internet, including its constituent routers and switches, end hosts, and data centers, using localized power sources such as solar or wind. These sources vary widely in the power they deliver on both short and long timescales [19]. We *do not* rely upon or assume the existence of a smart grid. While there are a number of projects making good progress on the development of a smart grid that integrates such diffuse and intermittent sources, we consider the more challenging case of operating networks that do not

---

*This work is supported by NSF grant CNS-0845349.

e-Energy 2012, May 9-11 2012, Madrid, Spain.

have consistent access to plentiful electricity. Even a smart grid will be unable to handle scenarios where the aggregate energy supply is less than demand or is too expensive to use consistently.

Next we describe the contours of the constraints we consider, and reasons why each constraint is fundamental.

The network edge is least power-constrained. Non-carbon energy sources are often diffuse. For reasons of resilience, efficiency, independence, and often politics, it may be difficult to produce such power centrally and then transmit it over long distances. Since edge nodes are typically deployed with less physical density than the core and data centers, the available energy per node at the edge may be greater than within the network.

Hard to build new infrastructure. While designing for such a set of challenging future scenarios, a natural response is to start from a clean slate. We argue that this is not possible in this particular instance much more so than other challenging future scenarios, because the challenge itself constrains the resources available to deploy new devices, systems, and infrastructure. That is, to build new infrastructure takes up-front energy—the energy to manufacture, deploy, operate, and maintain the devices, and as well as the financial resources to do so [22]. We contend that in an energy challenged world it will be far more efficient, in terms of energy and cost, to adapt today's infrastructure to the limits at hand.

Limited data and energy storage. A common approach to resolving problems of intermittency in either data or energy is to use buffers. Buffering approaches would likely rely upon non-volatile storage at data centers, such as in a delay-tolerant network. However, such buffering can be expensive. Similarly, non-hydrocarbon energy storage is neither dense nor inexpensive. As a result, we expect that little large-scale data or energy storage will be available.

Volatility. Coping with diffuse energy sources is further complicated by intermittency and the lack of data or energy storage. Historical experience has shown us that a volatile network is one that is hard to manage and is unforgiving for many protocols, including BGP [18], TCP [3], and others. Assuming constrained and intermittent power with limited storage, routers might power cycle frequently, causing frequent route fluctuations.

Demand exceeds supply. A common assumption in the design of networks today is that the supply of energy will be uninterrupted, limitless, and affordable enough that it seldom constrains the choice of whether to run systems at all. In the future we consider here, demand is likely to always exceed available energy supply.

A natural reaction is that these constraints are too restrictive, and that the future is unlikely to be constrained in precisely these ways. We agree, but our hope is that in designing an Internet that can cope with these constraints we build in a margin of resilience.

3. DESIGN

As the series of constraints discussed in Section 2 potentially manifest, how must the components of today's Internet architecture change? Here, we present the opposite of a conventional clean-slate Internet architecture design: clean-slate designs aim to create a radically different architecture using today's (energy) constraints, whereas our interest is in repurposing today's architecture and infrastructure for radically different energy constraints.

3.1 Design by Example

Although we do not have space to present a full design, we discuss a very concrete example to ground our discussion and to highlight challenges that arise. Our goal is not to present new technical solutions but rather to gather relevant ideas to meet the needs of an intermittent energy Internet. Our example is as follows: we wish to perform cross-continent data transfer—a server in California serves data to a client in Washington DC. A traceroute between these sites yields (layer 3) routers in Sacramento, San Jose, San Francisco, Kansas City, Chicago, and Washington DC. For simplicity we assume ideal renewable energy sources in each of these locations: the California and Washington DC nodes use solar and the Kansas City and Chicago nodes use wind [31].

We use historical weather data from Weather Underground for June 30th, 2011 to understand the renewable energy constraints the data transfer would have faced that day [30]. San Francisco and DC are sunny that day. However, Chicago has variable wind—from nearly calm to over 30mph—and Kansas City has low but constant wind. At the time Chicago's wind speed increases (around 9pm), San Francisco is getting dark and DC is already past dusk.

3.2 Components

Next we consider a redesign of today's Internet, keeping our example in mind as we do so.

3.2.1 Interdomain Routing

We examine the two key functions of routers: route computation/distribution and data-plane forwarding. Because we eschew clean-slate design, we do not consider the possibility of separating routing into a management layer [8].

What problems are encountered in each of these cases? Route distribution and computation is currently handled by BGP; even under good conditions BGP exhibits poor stability and is unlikely to fare well in a volatile network. On the day in question, Chicago is unlikely to be able to serve a significant amount of traffic until late in the day, and thus we have to use alternative routes. Also, the period of time Chicago has significant wind is on the order of an hour, which is not that long relative to the potential time of BGP convergence. Forwarding will have to be performed at a lower rate due to power constraints. Most notably, this scenario introduces a new dimension to routing decision-making: time.

If we were to replace conventional BGP with a variant that announces time-dependent routes, routing may become more static than today. This notion may seem counterintuitive at first, but to work within BGP's limited ability to cope with route flapping, we might be forced to pin routes and use hysteresis to prevent fluctuations from affecting route selection and causing update storms. Pinning would help alleviate the fluctuation that happens when hosts and routers unexpectedly and repeatedly fail using today's BGP. The goal would be to make BGP less reactive to temporary outages that occur during intermittent power failures.

For forwarding, we might consider power proportionality when parts of the route are power-constrained—e.g., switching off linecards when not needed [9]. Alternatively, if the hardware is such that transmission rate is proportional to power usage, as in the case of some wireless devices, we can decrease the transmission rate and still deliver the same data over a longer time period using less energy (since bit rate is a sub-linear function of transmit power according to Shannon's law [16]). Under more severe power constraints, we will simply have no choice but to shut off routers temporarily. If we were to do so, all forwarding must come to a stop and traffic will have to find other ways around.

3.2.2 Transport

Ensuring that the basic functions of data transport protocols (reliability, flow control, congestion control, service multiplexing) operate correctly is significantly complicated when end-to-end reachability itself is volatile. Operating TCP in challenging regimes is nothing new; research has explored using TCP in everything from sensor networks and sub-packet regimes to overloaded data centers and lossy wireless links. In a power-constrained network, transport protocols are likely to face high loss rates (often correlated rather than random), frequently interrupted flows, volatile bottleneck bandwidths, as well as sporadically available end systems.

In our example, we expect that for a substantial part of the day in question, connectivity between San Francisco and DC will be lossy and slow, requiring the use of either robust single-path transport protocols or multi-path transport via overlay IP tunnels. It seems appropriate for transport protocols in this scenario to begin and end communication quickly, with little overhead, and to be able to resume communication extremely quickly [25]. In addition, the traditional function of congestion control might disappear in this context as the fundamental tradeoff between protocol robustness and efficiency is exposed. New transport protocols that focus on conserving energy while still reliably delivering data rather than maximizing bandwidth usage will need to be developed.

3.2.3 Cloud

The cloud may be the most rapidly increasing energy consumer today, but faces challenges in a world of diffuse and intermittent power. If Chicago's power is limited for a large fraction of the day, running energy-intensive data centers is an untenable luxury. Any processing or access to data stored on those nodes would require careful scheduling to find times that all parties have sufficient power, which on the day in question only occurs for a brief period around 8:30pm Chicago time.

As with routing, one way to cope is to leverage power proportionality and simply shed load in order to decrease energy use. Under more severe constraints, a natural solution is to shut down servers as required and offload work to other machines [28]. In the event of anticipated power shortages, it would be possible to schedule batch/large jobs when power is available.

A more advanced approach might involve sharing the energy cost of cloud-based workloads by dynamically scaling the amount of work performed by the client vs. the server based on their available power. For example, if the client has significant solar power while using Google Docs and Google is near its power limit, Google can offload the dynamic features of the application to the client.

3.2.4 Addressing and Naming

In this scenario customer-provider relationships may become less clear as the policies available for use will need to be more flexible and the network service more redundant. As a result, more of what we think of as stub networks will require backup network service—how else would data have been routed from a host in Kansas City to a host in DC on the day in question?

Thus the meaning of fixed addresses in an intermittently powered network may change. If a server is flapping, should it receive a fixed address? What if a router is doing the same? There is significant debate over the meaning of addresses, the benefits of flat naming, and the historical conflation of location with identity. For convenience we think that fixed addressing is likely the best approach, but one could imagine the benefits of having a one-to-many anycast-like mapping for a broader set of addresses, extending the service model from today's to one in which an address refers to *the host that answers to this address and currently has power*.

3.2.5 Link/physical Technology

At the lowest level of communication, power availability affects the raw transmission of bits and signals over lambdas and copper. Power availability may have a wide range of effects depending upon the transmission technology in question, so we only consider the broad contours of these impacts on the link and physical layers.

At the link layer, many of the difficulties faced by challenged networks today are likely to arise. Thus the community's experience with highly-lossy wireless networks is beneficial when considering how to design a MAC protocol for a network in which recipient end points fluctuate in their availability. One option is to leverage broadcast-based mesh approaches such as ExOR [4].

Physical technologies also vary widely, but one commonality among many high-speed physical layer technologies is their use of circuits. Although these technologies operate on a smaller timescale than the timescales we anticipate for power fluctuation, they will nevertheless be more vulnerable to interruption than technologies that do not require circuit establishment.

3.2.6 Applications

Many application protocols require interactivity and fast responses in a way that may not be achievable in an energy constrained future. For example, HTTP, SMTP, IMAP, and other common protocols require significant interactivity during handshaking and client requests. Once a client has issued a request, it is free to either receive or send data as required. Worse still are many modern Ajax-based applications that require constant interactivity. However, without deploying modified applications, we might consider a stopgap solution of proxies designed to emulate the basic functionality of either end of a connection with minimal interactivity. Many transparent network proxies and de-duplication boxes have much of the functionality required to enable this today.

4. THE KEY TRADEOFF

As we saw in the previous section, the primary challenge we face is that point-to-point communication becomes much more difficult under constrained, intermittent power. The adaptations required for this future Internet will therefore encounter a key tradeoff between *resilience* and *efficiency*.

Resilience is often the opposite of efficiency; only a system that has fat to cut can survive when the fat is cut. In this context by resilience we mean that data can reach its destination in a predictable manner despite unpredictable changes in the network. By efficiency, we mean that the network itself is using as little energy as possible to perform the needed communications. Resilience requires overprovisioning the network to cope with unpredictability, but overprovisioning requires more resources, hence making the network less energy-efficient. Similarly, efficiency requires using the bare minimum resources to establish communication once resources become available, which means that communications are less resilient to network failure.

These two ends of the design spectrum have loose analogs in how traditional communication media operate. Thus, next we elaborate on how the Internet would look under the resilience-oriented model (akin to radio) and the efficiency-oriented model (akin to telephony). We believe there will be suitable applications for each model, and that a future Internet with intermittent power would likely have to be a combination of the two.

4.1 Resilience

The resilience-oriented model for optimizing an intermittent-energy Internet is loosely inspired by how radio stations broadcast content, and interested listeners tune in to seek the content they want to hear. To achieve resilience, this model would adopt techniques from delay-tolerant networking (DTN). Ideally, if there multiple paths between two end points, the communication mechanism will publish data along many or all paths simultaneously in case any one of them fails. Furthermore, even within a path there could be some unavoidable delays that require an intermediate node to store a copy of in-transit data onto stable storage (just in case one of the path links fail).

This is a publish-subscribe model similar to one proposed in studies on Information-Centric Networking [15, 17]. Since we are not interested in rearchitecting the Internet, we could adapt ICN principles using HTTP [24]. This model would be suited to applications such as bulk data transfer.

4.2 Efficiency

The efficiency-oriented model is loosely inspired by how telephone users only need the service for a brief period of time once they determine the appropriate time and conditions to communicate. This model achieves efficiency by having users pick the appropriate moment to initiate end-to-end communication, thereby reducing the total amount of energy required to transfer data in the system across all users. The primary challenge in this approach is developing a mechanism for determining both end point and in-network resource availability. To this end, we might leverage probabilistic routing and unconventional data sources such as weather models to schedule sessions when packet forwarding conditions are favorable, enabling interactive applications.

5. RELATED WORK

Reducing the energy use of servers [2], network protocols [23], networks [11], and data centers [21] has received significant attention over the last decade. However, in an energy-constrained future simply reducing energy consumption may not be enough, if even reduced energy demands still exceed the supply. Instead, systems will also have to cope with frequent and unpredictable fluctuations in available power from intermittent energy sources [29]. Recent work examines how data centers might cope, or take advantage of, intermittent renewable energy sources either by rapidly transitioning hardware components between a high-power active state and a low-power inactive state [14, 28] or by migrating data to data centers with plentiful power [1]. While a useful starting point, applying these techniques to the Internet may prove difficult, since a network path, unlike a data center, may include multiple locations with little power and mostly inactive components.

In addition to this body of recent work, delay-tolerant networking and developing regions research are good fits for our problem domain [5, 7]. However, since we do not consider a clean-slate approach, we find it hard to envision how a DTN or similar approach might be deployed at the scale required to meet the new needs of a downscaled Internet. One option would be to try to shoehorn the current IP architecture to make it delay tolerant without deploying real delay-tolerant functionality in the network.

6. CONCLUSIONS

We have just scratched the surface of the intellectual and practical challenges in building an intermittent-energy Internet architecture. Numerous questions and challenges remain, including pricing and scheduling of packets based upon energy use, coping with diurnal and seasonal effects, migrating physical infrastructure to regions with superior energy infrastructure, and exploring the broader architectural design space. All of these must be considered within the context of the central tradeoff of resilience and efficiency.

Coping with diffuse, intermittent, and expensive energy provides benefits regardless of how the future unfolds. In the best case, we enable a transition to an Internet that can run on renewable energy. In the worst case, we ensure the continued operation of the network, one that is robust to the uncertainties of a more chaotic future. We come away from our initial exploration optimistic given that our community appears well-equipped to address these constraints if we appropriately apply our knowledge, though knowledge is necessary but insufficient—we must follow it with action.

7. REFERENCES

[1] S. Akoush, R. Sohan, A. Rice, A. Moore, and A. Hopper. Free Lunch: Exploiting Renewable Energy for Computing. In *Proceedings of HotOS*, 2011.

[2] D. G. Andersen, J. Franklin, M. Kaminsky, A. Phanishayee, L. Tan, and V. Vasudevan. FAWN: A fast array of wimpy nodes. In *Proceedings of ACM SOSP*, 2009.

[3] H. Balakrishnan, R. H. Katz, and V. N. Padmanbhan. The effects of asymmetry on TCP performance. *Mobile Networks and Applications*, 4(3):219–241, 1999.

[4] S. Biswas and R. Morris. ExOR: opportunistic multi-hop routing for wireless networks. In *Proceedings of ACM SIGCOMM*, 2005.

[5] E. Brewer, M. Demmer, B. Du, M. Ho, M. Kam, S. Nedevschi, J. Pal, R. Patra, S. Surana, and K. Fall. The case for technology in developing regions. *IEEE Computer*, 38(6), 2005.

[6] J. J. Brown and S. Foucher. Egypt, a Classic Case of Rapid Net-Export Decline and a Look at Global Net Exports. *ASPO-USA*, Feb. 2011.

[7] K. Fall. A delay-tolerant network architecture for challenged internets. In *Proceedings of ACM SIGCOMM*, 2003.

[8] N. Feamster, H. Balakrishnan, J. Rexford, A. Shaikh, and J. Van Der Merwe. The case for separating routing from routers. In *Proceedings of the ACM SIGCOMM workshop on Future directions in network architecture*, 2004.

[9] W. Fisher, M. Suchara, and J. Rexford. Greening backbone networks: reducing energy consumption by shutting off cables in bundled links. In *Proceedings of the ACM SIGCOMM Workshop on Green networking*, 2010.

[10] J. Hamilton. Causes and Consequences of the Oil Shock of 2007–08. *Brookings Papers on Economic Activity*, 2009(1):215–261, 2009.

[11] B. Heller, S. Seetharaman, P. Mahadevan, Y. Yiakoumis, P. Sharma, S. Banerjee, and N. McKeown. ElasticTree: Saving Energy in Data Center Networks. In *Proceedings of USENIX/ACM NSDI*, 2010.

[12] R. Hirsch, R. Bezdek, and R. Wendling. Peaking of World Oil Production: Impacts, Mitigation, & Risk Management. *U.S. DOE NETL*, 2005.

[13] IPCC. Special Report on Renewable Energy Sources and Climate Change Mitigation. May 2011.

[14] D. Irwin, N. Sharma, and P. Shenoy. Towards Continuous Policy-driven Demand Response in Data Centers. In *Proceedings of the ACM SIGCOMM Workshop on Green Networking*, 2011.

[15] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard. Networking Named Content. In *Proceedings of ACM CoNEXT*, 2009.

[16] R. R. Kompella and A. C. Snoeren. Practical lazy scheduling in sensor networks. In *Proceedings of ACM SenSys*, 2003.

[17] T. Koponen, M. Chawla, B.-G. Chun, A. Ermolinskiy, K. H. Kim, S. Shenker, and I. Stoica. A Data-Oriented (and Beyond) Network Architecture. In *Proceedings of ACM SIGCOMM*, 2007.

[18] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian. Delayed Internet routing convergence. *IEEE/ACM ToN*, 9(3), June 2001.

[19] S. Low and M. Chandy. SMART Grid Strategy: Vision, Challenges and Solutions. 2010.

[20] B. McKibben. *Eaarth: Making a life on a tough new planet*. Henry Holt and Company, 2010.

[21] D. Meisner, C. Sadler, L. Barroso, W. Weber, and T. Wenisch. Power Management of Online Data-Intensive Services. In *Proceedings of ISCA*, 2011.

[22] T. Murphy. The Energy Trap. `http://physics.ucsd.edu/do-the-math/2011/10/the-energy-trap/`, 2011.

[23] S. Nedevschi, L. Popa, G. Iannaccone, S. Ratnasamy, and D. Wetherall. Reducing Network Energy Consumption via Sleeping and Rate-adaptation. In *Proceedings of USENIX/ACM NSDI*, 2008.

[24] L. Popa, A. Ghodsi, and I. Stoica. HTTP as the Narrow Waist of the Future Internet. In *Proceedings of the ACM SIGCOMM Hotnets Workshop*, 2010.

[25] S. Radhakrishnan, Y. Cheng, J. Chu, A. Jain, and B. Raghavan. TCP Fast Open. In *Proceedings of ACM CoNEXT*, 2011.

[26] B. Raghavan and J. Ma. Networking in the Long Emergency. In *Proceedings of the ACM SIGCOMM Workshop on Green Networking*, 2011.

[27] S. Schultz. Military study warns of a potentially drastic oil crisis. *Der Spiegel*, September 1, 2010.

[28] N. Sharma, S. Barker, D. Irwin, and P. Shenoy. Blink: Managing Server Clusters on Intermittent Power. In *Proceedings of ACM ASPLOS*, 2011.

[29] C. Stewart and K. Shen. Some Joules are More Precious than Others: Managing Renewable Energy in the Datacenter. In *Proceedings of HotPower*, 2009.

[30] Weather Underground. `http://www.wunderground.com`.

[31] Wind Energy Resource Atlas of the United States. National Renewable Energy Laboratory. `http://rredc.nrel.gov/wind/pubs/atlas/maps.html`.